On the Inducibility of Rooted Trees

by

Audace Amen Vioutou Dossou-Olory

Dissertation presented for the degree of

Doctor of Philosophy in Mathematics

in the Faculty of Science at Stellenbosch University

Supervisor: Prof. S. Wagner

December 2018

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2018 ....

Abstract

On the Inducibility of Rooted Trees

A. A. V. Dossou-Olory

Department of Mathematical Sciences
Mathematics Division
University of Stellenbosch
Private Bag X1, Matieland 7602, South Africa.

Dissertation: **PhD** (Mathematics)

December 2018

The density of appearances of a fixed tree in a larger tree is examined for rooted trees without vertices of outdegree 1 (also known as topological trees). Given a topological tree S with k leaves and an integer $n > k$, we are interested in finding the maximum and minimum number of isomorphic copies of S in an arbitrary n-leaf topological tree T. The problem becomes more relevant when n is sufficiently large. Then the goal becomes to determine the limit superior of the proportion of all subsets of k leaves of the set of leaves of T that induce a copy of S as the size of the tree T grows to infinity. This limiting maximum quantity is called the inducibility of S. We investigate the inducibility in topological trees at large: our major focus, however, is placed on bounded degree topological trees which we call d-ary trees—it is found that there is an explicit identity between the inducibility in topological trees and the inducibility in d-ary trees. We prove that the inducibility of every tree is strictly positive and also determine its explicit value for some special families of trees, namely stars, binary caterpillars, complete d-ary trees and more generally even d-ary trees. Further properties such as how much the inducibility differs asymptotically from the maximum density, are also studied. In particular, our results provide an affirmative answer to an existing conjecture on the inducibility of binary trees. We also solve (at least approximately) another open question concerning the inducibility of a binary tree with five leaves—part of this is done by means of an algorithmic approach. Finally, we consider the problem of finding the asymptotic minimum number of copies of a d-ary tree. For the minimum, the situation is quite different from that of the maximum; we show that, in the degree-restricted context, the limit inferior of the proportion of all subsets of k leaves of the set of leaves of T that induce a copy of S as the size of T tends to infinity is positive for binary caterpillars (a binary tree with the property that a rooted path remains upon removal of all leaves) only. This allows us to derive an explicit lower bound on this limiting quantity.

Uittreksel

Oor die Indusibiliteit van Gewortelde Bome

("On the Inducibility of Rooted Trees")

A. A. V. Dossou-Olory

Departement Wiskundige Wetenskappe,
Wiskunde Afdeling
Universiteit van Stellenbosch
Privaatsak X1, Matieland 7602, Suid Afrika.

Proefskrif: **PhD** (Wiskunde)

Desember 2018

Die digtheid van verskynings van 'n vaste boom in 'n groter boom word ondersoek vir wortelbome sonder nodusse van uitgangsgraad 1 (ook bekend as topologiese bome). Gegewe 'n topologiese boom S met k blare en 'n heelgetal $n > k$, stel ons belang daarin om die maksimum en minimum aantal isomorfe kopieë van S in 'n willekeurige n-blaar topologiese boom T te bepaal. Die probleem word meer relevant wanneer n groot genoeg is. Dan word die doel om die bolimiet te bepaal van die verhouding van alle subversamelings van k blare van T wat 'n kopie van S voorstel, as die grootte van die boom T ba oneindig strewe. Hierdie maksimum word in die limiet die indusibiliteit van S genoem. Ons ondersoek die indusibiliteit in topologiese bome oor die algemeen: ons belangrikste fokus word egter op topologiese bome met beperkte grade geplaas wat ons d-êre bome noem — dit word gevind dat daar 'n eksplisiete verband bestaan tussen die indusibiliteit in topologiese bome en die indusibiliteit in d-êre bome.

Ons bewys dat die indusibiliteit van elke boom streng positief is en bepaal ook die eksplisiete waarde daarvan vir sommige spesiale bome, naamlik sterre, binêre ruspes, volledige d-êre bome en meer algemeen ewe d-êre bome. Verdere eienskappe soos hoeveel die indusibiliteit asimptoties van die maksimum digtheid verskil, word ook bestudeer. In die besonder lewer ons resultate ’n bevestigende antwoord op ’n bestaande vermoede oor die indusibiliteit van binêre bome. Ons beantwoord ook (ten minste met ’n benadering) ’n oop vraag oor die indusibiliteit van ’n sekere binêre boom met vyf blare — ’n deel hiervan word deur middel van ’n algoritmiese benadering gedoen. Laastens beskou ons die probleem om die asimptotiese minimum aantal kopieë van ’n d-êre boom te vind. Vir die minimum is die situasie heel anders as dié vir die maksimum; ons wys dat in die graadbeperkte konteks die onderlimiet van die verhouding van alle subversamelings van k blare van T wat ’n kopie van S voorstel as die grootte van T na oneindig strewe net vir binêre ruspes (’n binêre boom met die eienskap dat ’n gewortelde pad oorbly as alle blare verwyder word) positief is. Dit stel ons in staat om ’n eksplisiete ondergrens vir hierdie limiethoeveelheid af te lei.

Acknowledgements

"La vie n'est bonne qu'à deux choses: découvrir les mathématiques et enseigner les mathématiques." – S. D. Poisson (1781-1840)

I convey my heartfelt gratitude to my supervisor, Prof. Stephan Wagner who gets to shape and inspire my mind throughout the PhD program; I am especially grateful for his patience with all sorts of questions and making me believe in myself, that I could always do things better. I am also honoured that Professors Hua Wang (Georgia Southern University, USA), Peter Dankelmann (University of Johannesburg, ZA) and Marcel Wild (Stellenbosch University, ZA) have devoted time and energy into reviewing this manuscript—I am fully indebted to them.

I acknowledge AIMS-South Africa for providing me with a partial bursary towards this research, and Prof. S. Wagner for offering me not only complementary funding towards the work but also many other wonderful supports that allowed me to internationalise our research and forge a network with other scientists. I am likewise indebted to Stellenbosch University for offering a conducive environment and also providing me with the post-graduate merit bursary.

My deepest thanks go out to my parents in Benin for their unflinching love and prayers, particularly my mum Eléonore Nonvidé, my dad Pascal Dossou-Olory, my pépé Bienvenu Olory, and all my sibling sisters and brothers. Specials thanks to my well-wisher, Triphène Alladayè for her understanding and patience while being away for studies.

Dedications

A Dieu Tout-Puissant, Éternel des Armées, Alpha et Oméga, Chemin, Vérité et Vie, for all wonderful accelerations in my life.

In memory of my late brother, Princio Dossou-Olory (04/07/1990–05/07/2016), <the pain of your earlier passing, I shall forever carry it with me>.

"For there is hope of a tree, if it be cut down, that it will sprout again, and that the tender branch thereof will not cease."- ***Job 14:7***

Contents

List of Figures

List of Tables

Nomenclature

Symbols and Definitions

$\delta_{m,n}$	The Kronecker delta function; it takes on value 1 if $m = n$, and 0 otherwise
$\|T\|$	The number of leaves of the tree T
$L(T)$	The set of leaves of the tree T
$c(S,T)$	The number of copies of the tree S in the tree T
$c_l(S,T)$	The number of copies of the tree S in the tree T in which the leaf l of T is involved
$\gamma(S,T)$	The density of the tree S in the tree T
$I_d(S)$	The inducibility of the tree S in d-ary trees
$i_d(S)$	The inducibility of the tree S in strictly d-ary trees
$J(S)$	The inducibility of the tree S in topological trees
C_k	The star with k leaves
F_k^d	The d-ary caterpillar with k leaves
CD_h^d	The complete d-ary tree with height h
N_n	The number of topological trees with n leaves
N_n^d	The number of d-ary trees with n leaves

$\mathcal{P}^d(n)$	The set of partitions of n of length at most d but at least 2
$\lfloor y \rfloor$	The greatest integer less than or equal to y
$\lceil y \rceil$	The smallest integer greater than or equal to y
$\mathcal{F}(S_1; S_2)$	The tree obtained by appending the root of the tree S_2 to every leaf of the tree S_1
$V_{d,k}$	The set of all d-tuples of nonnegative integers that sum to k and whose largest part is less than k

Chapter 1

General introduction

Given a rooted tree S with k leaves and an integer $n \geq k$, what can be said about the extremal numbers of isomorphic copies of S among all rooted trees with n leaves? As the question becomes substantially more interesting when n gets large, we shall provide an asymptotic answer (as $n \to \infty$) in this thesis. The asymptotic maximum number of copies of S is captured by the graph invariant called *inducibility* of S.

1.1 Objectives

The inducibility of a rooted tree with k leaves is a novel graph-theoretic concept that has been put forward recently in [1] to deal with a typical problem stemming from the phylogenetics context in mathematical biology. In a compact way, the problem that is addressed in [1] can be stated as *maximising the asymptotic density* of appearances of 'small' rooted binary trees in rooted binary trees with 'large' number of leaves. The present work is aimed at extending this newly introduced concept from binary trees to rooted trees without vertices of outdegree 1. We primarily investigate the inducibility in rooted trees with bounded degrees and no vertices of outdegree 1. Later, we relax the degree restriction and show that there is a striking connection between the inducibility in trees with bounded degrees and the inducibility in trees (without vertices of outdegree 1) at large. Finally, we consider the opposite problem concerning the minimum number of isomorphism copies of a fixed rooted tree with k leaves among all rooted trees with $n \geq k$ leaves.

1.2 An overview on prior work

The problem of maximising the density of graphs in larger graphs is a key concept that has been explored starting from the work [2] of Martin C. Golumbic and Nicholas J. Pippenger, and has ever since been of interest to graph theorists. In their article [2] published in the Journal of Combinatorial Theory in 1975, Golumbic and Pippenger initiated the subject of the inducibility for simple graphs (i.e., nonoriented graphs without multiple edges and loops) in the following manner: for finite and simple graphs G and H with k and n vertices, respectively, let $\mathcal{I}(G, H)$ be the number of distinct subgraphs induced by k distinct vertices of H which are isomorphic to G; the quantity

$$I(G) := \lim_{n \to \infty} \left(\max_{|H|=n} \mathcal{I}(G, H) / \binom{n}{k} \right),$$

where the maximum runs over all finite and simple graphs on $n \geq k$ vertices, is called the *inducibility* of the graph G. In this setting, it is the asymptotic behaviour of the maximum number of appearances of G as a subgraph in an arbitrary n-vertex graph as $n \to \infty$, which is captured by the inducibility $I(G)$. The authors of [2] furnished some properties regarding $I(G)$, and calculated the precise inducibility of the complete bipartite graphs $K_{k,k}$ and $K_{k,1+k}$ (as well as their equivalent graph complements) for every $k \geq 1$.

After the original work [2], there have been numerous investigations on $I(G)$ in special cases. Indeed, the maximum number of induced subgraphs of H isomorphic to the complete bipartite graph $K_{k,k}$ has been studied in a 1986 paper [3] by Bollobás, Nara and Tachibana. In particular, the value of $I(K_{k,k})$ was rediscovered by Bollobás et al. A great deal of work (see [4; 5; 6]) followed [3] afterwards. As such, Brown and Sidorenko [5] computed $I(K_{k,k+l})$ explicitly for all $l \geq 1$ but under the restriction $k \geq l(l-1)/2$. For $k < l(l-1)/2$, they stated the result as a function of the maximum over $[0, 1]$ of a certain polynomial in a single variable. In 2014, James Hirst [7] determined, employing Razborov's flag algebra [1] method and semi-definite programming techniques [2], the inducibility of two 4-vertex graphs, namely the complete tripartite graph $K_{1,1,2}$ and the so-called paw graph (graph constructed from a triangle by appending a pendant edge). The concept of inducibility is still gaining consideration from several research groups; refer to [15] and [16] for some recent results on so-called blow-up of graphs and graphs on four vertices, respectively. The language of flag algebra was also employed recently by Balogh, Hu, Lidický and Pfender [17] to derive the inducibility of the cycle on five vertices, thereby settling a particular case of a conjecture formulated by Golumbic and Pippenger in [2].

Let us mention that there has also been interest in oriented graphs (with no multiple edges and no loops). For instance, in 2011, Sperfeld [18] explored, by means of flag algebra, the inducibility of (monodirected) graphs with at most four vertices. Three years later, Huang [19] determined, by means of a different approach, the maximum induced proportion of directed star graphs and also worked on some related problems.

Inducibility of Trees: One of the best investigated classes of graphs is the class of trees. This is due to their numerous applications in all of science. For example, biologists and in particular geneticists use trees in their models—refer to [20] for a recent work in theoretical biology; rooted trees with maximum degree at most four are DNA graphs; computer scientists are commonly using trees in data collection, processing, and analysis of algorithms—[21; 22]; rooted trees have found uses in the analysis of the order conditions of Runge-Kutta methods—[23; 24]; in organic as well as quantum chemistry, trees function as molecular graphs of acyclic organic molecules, portraying a large body of biochemical reactions such as molecules and molecular compounds—[25; 26]; trees are formal object of study (from a purely mathematical view) for combinatorists—[27; 28], etc.

Due to the prevalence of trees in many situations, the study of the inducibility of trees emerged in 2016, and has been addressed in two different settings. In [29], Bubeck and Linial investigated what they called *k-profile*

[1] Flag algebra is a modern theory initiated by Alexander Razborov for solving extremal problems of a certain sort (typically density of a submodel in a "large" structure) in combinatorics and graph theory; see [8; 9; 10; 11] for resources on this theory.

[2] These techniques are used for instance in [12; 13]; see also Chapter 6 of [14].

of trees, which is the vector of induced proportions of trees with k vertices classified according to their isomorphism type. At the end of their paper, Bubeck and Linial briefly defined a notion of inducibility of trees in this specific situation. Analogously, originally motivated by a question from phylogenetics (a branch of mathematical biology), namely that of finding the smallest possible number of crossing pairs of matching edges in tanglegrams [3], Czabarka, Székely and Wagner [1] introduced a further variant of the concept by studying the inducibility of a fixed rooted binary tree in binary trees. Loosely speaking, the inducibility of a tree is a measure of the maximum frequency at which the tree can be isomorphically embedded in a very 'large' tree.

For rooted binary trees B, T with $|B|$ and $|T|$ leaves respectively, let us denote by $c(B, T)$ the number of subtrees induced by $|B|$ distinct leaves of T which are homeomorphically irreducible [4] to a tree isomorphic to B. The *inducibility* $i(B)$ of B, as defined and studied in [1], is given by

$$i(B) = \limsup_{n \to \infty} \left(\max_{|T|=n} c(B, T) / \binom{n}{|B|} \right),$$

where the maximum is taken over all rooted binary trees with $n \geq |B|$ leaves. A natural related problem for further study is the inducibility of rooted trees with bounded degrees and no vertices of degree 2, except possibly the root. The special case of binary trees is also known as phylogenetic trees that are used in biology to describe, for instance, how entities (such as species, populations, organisms) are evolutionarily linked [31; 33]. Of a purely mathematical interest, we shall also investigate the inducibility in rooted trees without restriction on the degree sequence. As a particular case of our results, we shall provide an affirmative answer to a conjecture formulated by Czabarka, Székely and Wagner in [1]. In addition, we shall give an approximative answer to another question left open in [1].

[3] Tanglegrams are a special kind of graphs which are objects of study in evolutionary biology [30; 31; 32].

[4] A formal definition is given in the next chapter.

1.3 Layout of the thesis

The outline of the dissertation may be sketched as follows: Chapter 2 is of a preparatory nature – it covers some important preliminary details such as selected terminologies, tools and first established results that are used in the thesis. Chapters 3 to 7 form our main contribution; they actually represent papers that have been submitted/accepted for publication [34; 35; 36; 37; 38] in journals. A short version of Section 3.4 of Chapter 3 was presented in a poster session and flash at the *5th Heidelberg Laureate Forum* (Germany, September 24-29, 2017).

In Chapter 3, we examine the inducibility in (rooted) d-ary trees; how much it differs asymptotically from the maximum density; and its explicit value in two special families of d-ary trees, namely stars and binary caterpillars. We also discuss the characterisation of the d-ary trees that have the maximal inducibility. In Chapter 4, we investigate the inducibility in (rooted) topological trees at large, and also find the maximal trees in this situation. We show an explicit relation between the inducibility in topological trees and its d-ary counterpart. We also look at various lower bounds involving solely the number of leaves of the tree. Gradually, we extend the work by deriving in Chapter 5 the precise inducibility of every even d-ary tree in d-ary trees, furnishing bounds on the inducibility of an arbitrary balanced d-ary tree, and also considering further structural restrictions on the tree. Another main result of Chapter 5 is a general inequality between the inducibilities of a fixed d-ary tree and its branches. We also prove that the inequality holds with equality for infinitely many d-ary trees.

In Chapter 6, we focus on topological trees with at most five leaves, and provide further lower and upper bounds on their inducibilities. The upper bound is established using an algorithmic approach. In particular, one of our main results is an approximative answer to an existing question on the inducibility of a 5-leaf binary tree. Our final Chapter 7 is concerned with the minimisation part of the work. We provide a complete solution to the problem of finding the asymptotic minimum density of a binary caterpillar in d-ary trees. We also derive an explicit formula for the number of copies of a d-ary caterpillar in a complete d-ary tree of arbitrary height.

The work also emphasises future directions of research on the present topic—these suggestions are incorporated throughout, in chapters.

Chapter 2

Terminologies and preliminary statements

The purpose of this chapter is to fix in one place the main vocabulary and present some ingredients needed for most of our analysis. Furthermore, we give a formal definition of the inducibility and also provide some basic results on the inducibility as a running example. New terms will be defined in some other chapters when their time has come.

2.1 Main terminologies

We shall not recall graph-theoretic terms; for a thorough resource on the basic concepts of graph theory, we refer to West et al. [39], and Harary and Palmer [40]. We start with some conceptual facts on trees and their structure.

2.1.1 Basics of trees

A *tree* is a simple (undirected, no multiple edges, no loops) connected acyclic graph. Equivalently, any two vertices of a tree are connected by a unique and simple path. A degree 1 vertex in a tree is called a *leaf*. A *rooted* tree is a tree in which one of its vertices is designated to serve as and to be called the root—sometimes, the root is the topmost vertex in a graphical representation of the tree and each edge is then implicitly directed away from the root [1]. The *height* of a rooted tree is defined as the length of a longest path from the root to a leaf (see Figure 2.1 for a rooted tree of height 3).

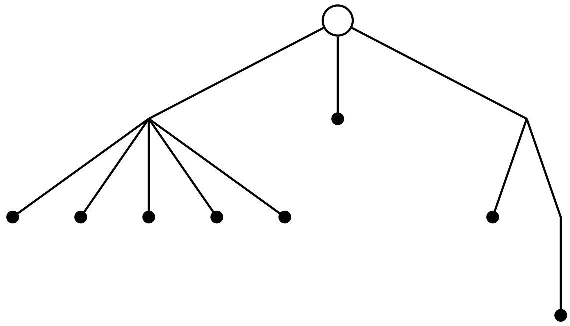


Figure 2.1: A rooted tree T_1 of height 3 with 8 leaves.

Rooted trees have an evolutionary structure: if vertex u immediately precedes vertex v on the single path from the root to v, then we say that u is the *parent* of v and v is a *child* of u. In particular, a leaf of a rooted tree is any vertex having no children. The non-leaves (including the root) are called *internal* vertices. The tree that consists of only one vertex will be treated as both leaf and root. The children of the root r of a rooted tree T are called the roots of the branches of T, so that the subtree induced by a child u of r and all its descendants u_i (i.e., u is on the unique path from the root to every u_i) is called a *branch* of T. In other words, the branches of T are the (connected) components that remain when the root r (together with all its incident edges) is suppressed. Thus, the branches of a rooted tree (with root r) are themselves rooted trees (endowed with their natural roots) and the number of branches is just the total number of children of r (i.e., the degree of r). In particular, the number of leaves in a non-trivial (at least two leaves) rooted tree is simply the sum of the number of leaves of its branches.

The root of a tree is of a particular interest for us: we shall think of two rooted trees as being *isomorphic* if there is a graph isomorphism between them (preserving adjacency) that maps the root of the one to the root of the other. In simple language, one tree can be obtained from the other by a

[1]This convention is adopted throughout the present work.

finite sequence of rotating and/or shifting edges around in such a way that root and edge structures are preserved. We then regard two rooted trees as the same/identical if and only if they are isomorphic. Note that viewing two isomorphic trees as the same also means that we do not distinguish between the various *embeddings* of a tree in the plane. For example, the tree depicted in Figure 2.2 is isomorphic to the one shown in Figure 2.1. Hence, we consider them as the same, i.e., $T_1 = T_2$.

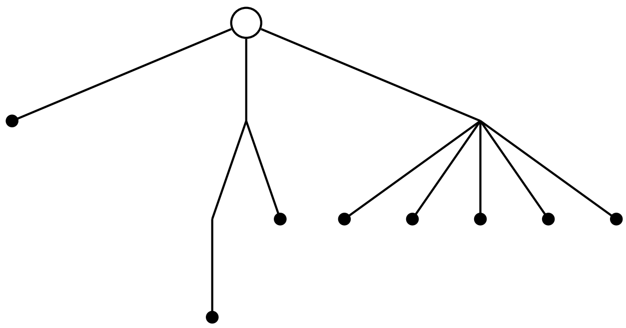


Figure 2.2: A rooted tree T_2 isomorphic to the rooted tree in Figure 2.1.

Clearly, there are more isomorphism types of rooted trees than there are of unrooted trees. We mention that a linear time algorithm in the number of vertices is available for testing the isomorphism of two rooted trees, see [22] for instance.

2.1.2 Topological trees and leaf-induced subtrees

If v is a vertex of a rooted tree, then the number of children of v is called the *outdegree* of v. A tree that does not contain a vertex of degree 2 is called a *topological* tree (see e.g. Bergeron, Labelle and Leroux [41], or Allman and Rhodes [42]), or also *series-reduced* or *homeomorphically irreducible* tree. Based on this definition, we shall call any rooted tree in which every vertex has outdegree at least 2 a (rooted) topological tree. For technical reasons, we also define the tree that has only one vertex as a topological tree.

The next question is typically enumeration. It is not hard to give a list of topological trees with a small number of leaves. For example, Figure 2.3 displays all the topological trees with fewer than five leaves. The following formula[2] given by Genitrini [43] computes the number N_n of topological trees with n leaves: $N_0 = 0,\ N_1 = 1$ and for $n > 1$,

$$N_n = \frac{1}{n} \sum_{\substack{m|n \\ m<n}} m \cdot N_m \; + \; \frac{2}{n}\Big(\sum_{j=1}^{n-1} j \cdot N_j \sum_{m=1}^{\lfloor \frac{n-1}{j} \rfloor} N_{n-j\cdot m} - \frac{1}{2}\delta_{n-j\cdot m,1}\Big),$$

where the first sum ranges over all positive divisors of n other than n and $\delta_{j,k}$ is the Kronecker delta function. The counting sequence N_n for the number of n-leaf topological trees ($n \geq 1$) starts

$$1, 1, 2, 5, 12, 33, 90, 261, 766, 2312, 7068, \ldots,$$

see `A000669` in Sloane's Online Encyclopedia of Integer Sequences [44] for more information.

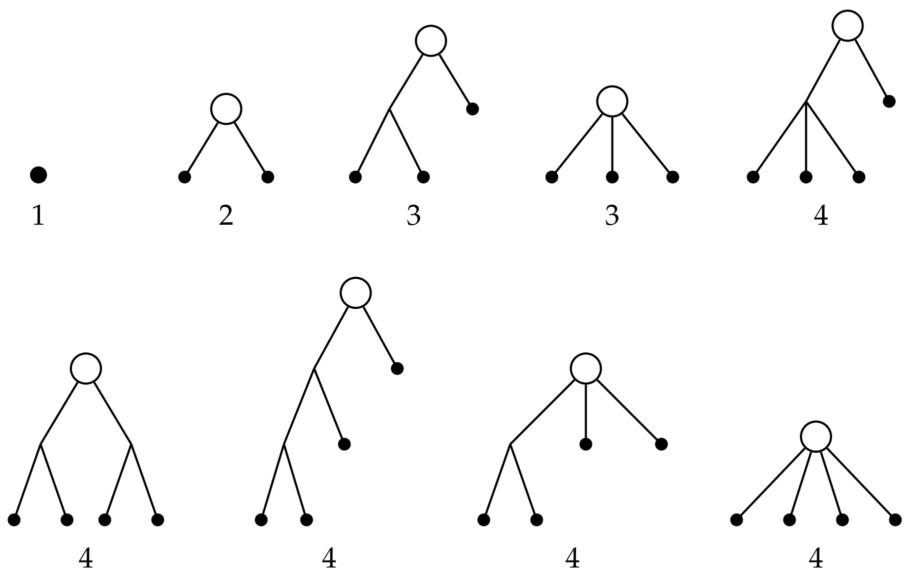


Figure 2.3: All the topological trees with fewer than five leaves.

Note that it is important to not have vertices of outdegree 1 in our context, because we are collecting trees according to the number of leaves, and there are infinitely many trees with the same number of leaves (e.g. all paths; more generally, one can subdivide any set of edges in a given tree). So this restriction is naturally needed. Another motivation is that rooted trees

[2]The formula was derived by means of a generating function approach.

where only the root may have degree 2 are the relevant trees for biological applications.

Every rooted tree can be (*homeomorphically*) reduced to a topological tree by suppressing all degree 2 vertices from the original tree. If L is a subset of the leaf set of a topological tree T, then L induces another topological tree which is obtained through the following operation: first extract the minimal subtree of T containing all the leaves in L, and then suppress (if any) all vertices whose outdegree is 1. Such a subtree will be referred to as a *leaf-induced subtree* of T. It has a root in a natural way which we define as the most recent common ancestor shared by the leaves in L. It follows from the definition that there is a unique leaf-induced subtree with k leaves of an arbitrary n-leaf topological tree ($n \geq k$) for each of the $\binom{n}{k}$ possible choices of k distinct leaves. Figure 2.5 presents a step-by-step illustration of the operation that gives a leaf-induced subtree of the topological tree shown in Figure 2.4. In this specific example, the root of the leaf-induced subtree coincides with the root of the original tree (which is the only ancestor that the leaves l_1, l_2, l_3, l_4, l_5 have in common).

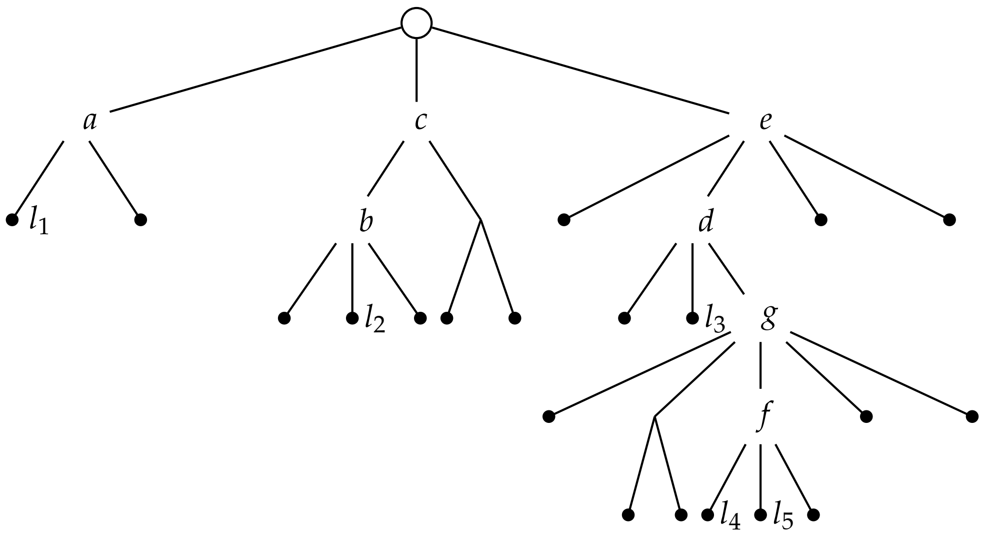
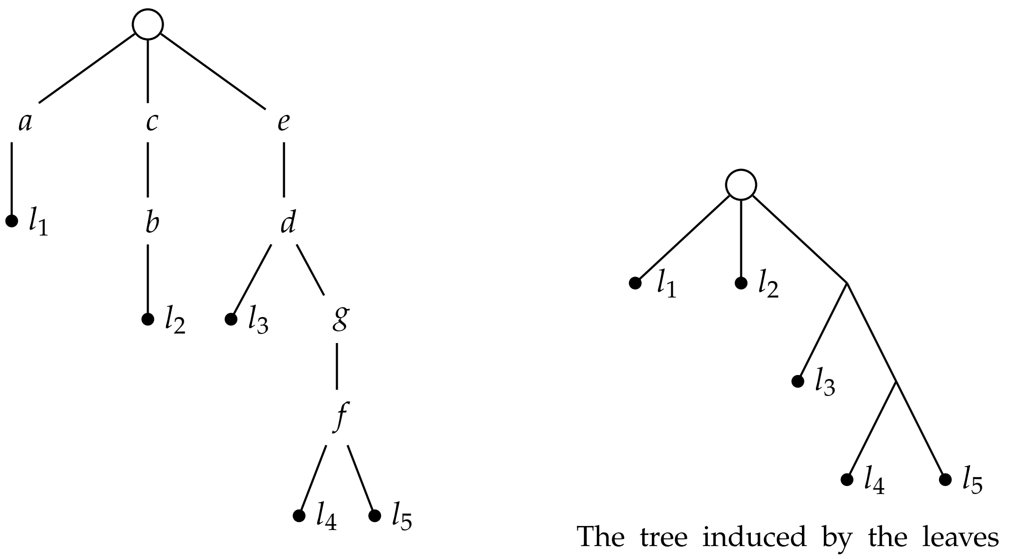


Figure 2.4: A topological tree and five of its leaves: $\{l_1, l_2, l_3, l_4, l_5\}$.

We shall write $|T|$ for the number of leaves of a topological tree T.

Let $d \geq 2$ be an arbitrary but fixed positive integer. A topological tree will be called a *d-ary tree* if each of its non-leaf vertices has no more than d



The minimal tree containing the leaves l_1, l_2, l_3, l_4, l_5.

The tree induced by the leaves l_1, l_2, l_3, l_4, l_5 after erasing the vertices a, b, c, e, g.

Figure 2.5: The minimal tree containing the leaves l_1, l_2, l_3, l_4, l_5 of the tree shown in Figure 2.4 and the leaf-induced subtree.

children. The tree that has only one vertex is also called a d-ary tree. We shall simply refer to a 2-ary tree as a *binary* tree, and a 3-ary tree will be called a *ternary* tree.

The next question is typically enumeration. One can derive a general recursive formula for counting the number of nonisomorphic d-ary trees with a given number of leaves:

Proposition 2.1.2.1. Let $d \geq 2$ be an arbitrary but fixed positive integer. If N_n^d denotes the number of nonisomorphic d-ary trees with n leaves, then for $n \geq 2$, we have

$$N_n^d = \sum_{(\underbrace{k_1 = \cdots = k_1}_{\alpha_1 \text{ times}} > \underbrace{k_2 = \cdots = k_2}_{\alpha_2 \text{ times}} > \cdots > \underbrace{k_m = \cdots = k_m}_{\alpha_m \text{ times}}) \in \mathcal{P}^d(n)} \prod_{i=1}^{m} \binom{N_{k_i}^d + \alpha_i - 1}{\alpha_i}, \tag{2.1.1}$$

where $\mathcal{P}^d(n)$ represents the set of all partitions of n of length at least 2 but at most d.

Proof. We regard a partition (of length l) of n as a nonincreasing sequence $n_1 \geq n_2 \geq \cdots \geq n_l$ of positive integers such that $\sum_{i=1}^{l} n_i = n$. For arbitrary d-ary trees $D_1, D_2, \ldots, D_l$ such that $\sum_{i=1}^{l} |D_i| = n$, we can form an n-leaf d-ary tree by attaching the roots of the D_i to a (new) common vertex. That way, given an element

$$\big(\underbrace{k_1 = \cdots = k_1}_{\alpha_1 \text{ times}} > \underbrace{k_2 = \cdots = k_2}_{\alpha_2 \text{ times}} > \cdots > \underbrace{k_m = \cdots = k_m}_{\alpha_m \text{ times}}\big)$$

of $\mathcal{P}^d(n)$, the α_i represent the multiplicities (in terms of number of leaves) of the branches of the associated n-leaf tree. Thus, choosing each of the m different branches to build the n-leaf tree is equivalent to choosing α_i d-ary trees with k_i leaves each (with the possibility of choosing the same tree more than once) among the $N_{k_i}^d$ available d-ary trees, for every $i \in \{1, 2, \ldots, m\}$. This proves the required identity in the proposition. □

The recursion (2.1.1) does not have an explicit (closed) formula for general d. Table 2.1 indicates the first few values of N_n^d obtained by means of a computer program.

Table 2.1: The number N_n^d of d-ary trees with n leaves.

Values of N_n^d for $d = 3, 4, 5$.													
n	1	2	3	4	5	6	7	8	9	10	11	12	13
$d = 3$	1	1	2	4	9	23	58	156	426	1194	3393	9802	28601
$d = 4$	1	1	2	5	11	30	80	228	656	1945	5835	17808	54881
$d = 5$	1	1	2	5	12	32	87	251	733	2201	6696	20705	64681

An equivalent recursive formula is known for the special case $d = 2$ (*Wedderburn-Etherington numbers*). Starting with $N_1^2 = 1$, we have

$$N_{2n}^2 = \frac{1}{2}\Big(N_n^2 + \sum_{i=1}^{2n-1} N_i^2 \cdot N_{2n-i}^2\Big), \; N_{2n+1}^2 = \frac{1}{2}\Big(\sum_{i=1}^{2n} N_i^2 \cdot N_{2n+1-i}^2\Big),$$

which can also be explained as follows: for odd number of leaves, every binary tree is counted exactly twice (swapping the left and right branches) in this way, whereas for even number of leaves, a binary tree for which the two branches have the same number of leaves is counted exactly once in this way. We provide in Table 2.2 the first values of N_n^2.

Table 2.2: Number of binary trees with n leaves.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N_n^2	1	1	1	2	3	6	11	23	46	98	207	451	983	2179	4850

For further information about the counting sequence $(N_n^2)_{n\geq 1}$, we refer to `A001190` in [44]. We remark that the counting sequence for the general case N_n^d appeared very recently in [44] for the cases $3 \leq d \leq 10$ as `A268172`, `A292210`, `A292211` through `A292216`, respectively.

A d-ary tree in which every vertex has exactly 0 or d children will be called a *strictly* d-ary tree. So all the branches of a strictly d-ary tree are themselves strictly d-ary trees; see Figure 2.6 for an example of a strictly d-ary tree.

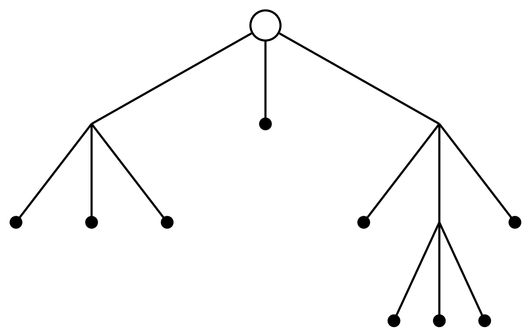


Figure 2.6: A strictly ternary tree with height 3 and 9 leaves.

The restriction of $\mathcal{P}^d(n)$ to partitions of length exactly d in Proposition 2.1.2.1 yields the number of nonisomorphic strictly d-ary trees with $n \geq 2$ leaves; of course, this only exists for every $n \equiv 1 \bmod (d-1)$ (which is easily seen by induction on n) for every given $d \geq 2$.

A *complete* d-ary tree is a strictly d-ary tree in which all the leaves (d^h in total) are at the same distance h from the root. The branches of a complete d-ary tree are themselves complete d-ary trees. The complete d-ary trees of height 1 will also be referred to as *stars*. We shall denote the complete d-ary tree of height h by CD_h^d. See Figure 2.7 for the complete 4-ary tree of height 2.

Anticipating on what we shall see later, complete d-ary trees can be thought of as a family of d-ary trees that contains 'almost' every d-ary tree. We shall say more about this family of trees in a later chapter.

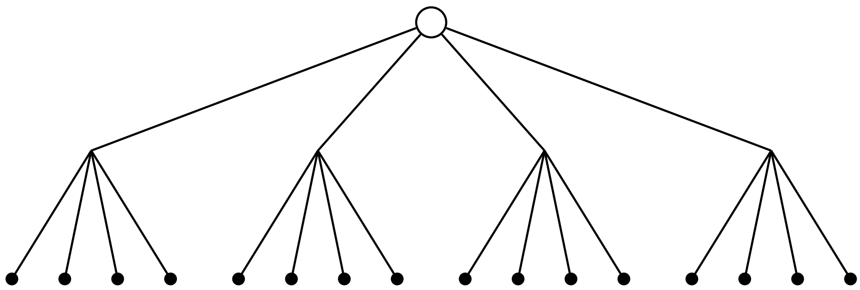


Figure 2.7: The complete 4-ary tree of height 2.

Before continuing, we wish to supply a first result which motivates our study. Denote by C_3 the 3-leaf strictly ternary tree (the star with 3 leaves). We are going to compute the maximum proportion, as $n \to \infty$, of leaf-induced subtrees with three leaves isomorphic to C_3 that can occur in a strictly ternary tree T with n leaves. We call this quantity the *limiting maximum density* of C_3 in strictly ternary trees.

To be precise, we are going to calculate the greatest accumulation point of the sequence

$$\left(\max_{\substack{|T|=n \\ T \text{ strictly ternary tree}}} \frac{c(C_3, T)}{\binom{n}{3}} \right)_{n \geq 1} .$$

Proposition 2.1.2.2. The limiting maximum density of C_3 in strictly ternary trees is $1/4$.

Proof. We prove the statement using a well-known technique in calculus. For a strictly ternary tree T with n leaves, we show that

$$c(C_3, T) \leq \frac{n(n-1)(n+1)}{24} ,$$

and moreover, the inequality holds with equality if and only if T is a complete ternary tree.

Since for $|T| = n \leq 2$, there are obviously no copies of C_3 in T, and the bound on $c(C_3, T)$ is a nonnegative quantity, we deduce that the inequality holds for $n \leq 2$. We may then assume that the tree T has more than two leaves and reason by induction on $|T| = n$. For convenience, let us use the abbreviation

$$P(n) := \frac{n(n-1)(n+1)}{24} .$$

For the induction step, consider a strictly ternary tree T with $n \geq 3$ leaves and denote by T_1, T_2, T_3 its three branches. Suppose that $|T_1| = k_1, |T_2| = k_2$ and $|T_3| = k_3$ (so $k_1 + k_2 + k_3 = |T| = n$). We note that there are only three possible scenarios that can happen for a subset of three leaves of T:

- Case 1: All three leaves belong to the same branch of T. The total number of these subsets of leaves of T that induce C_3 is

$$c(C_3, T_1) + c(C_3, T_2) + c(C_3, T_3)\,.$$

- Case 2: Two of the leaves belong to one branch of T, the third leaf to one of the two other branches of T. But any such set of three leaves induces a tree whose root has exactly two children and so cannot be isomorphic to C_3, giving no copies of C_3.

- Case 3: Each of the branches of T contains one of the three leaves. In this case, the three leaves always induce the tree C_3 (consisting of a root and three leaves attached to it), yielding $k_1 \cdot k_2 \cdot k_3$ copies of C_3.

Putting the three cases together, we find the recursion

$$c(C_3, T) = c(C_3, T_1) + c(C_3, T_2) + c(C_3, T_3) + k_1 \cdot k_2(n - k_1 - k_2)\,,$$

which gives the total number of copies of C_3 in T. Now using the induction hypothesis, we get the following inequality:

$$c(C_3, T) \leq P(k_1) + P(k_2) + P(n - k_1 - k_2) + k_1 \cdot k_2(n - k_1 - k_2)\,.$$

Consider the function

$$f(k_1, k_2) := P(k_1) + P(k_2) + P(n - k_1 - k_2) + k_1 \cdot k_2(n - k_1 - k_2)$$

in the two variables k_1 and k_2. One can see that f is a continuous function in the closed and bounded domain given by the inequalities $1 \leq k_1, k_2 \leq n - 2$. First, we solve the system $\nabla f(k_1, k_2) = (0, 0)$ of first-order partial derivatives of f, and we find

$$\Big(\frac{n}{3}, \frac{n}{3}\Big), \Big(\frac{7n}{9}, \frac{n}{9}\Big), \Big(\frac{n}{9}, \frac{7n}{9}\Big), \Big(\frac{n}{9}, \frac{n}{9}\Big)$$

as the stationary points of f. Next, for every stationary point (k_1^*, k_2^*), we compute the determinant

$$\det H(k_1^*, k_2^*) = \frac{(n-9k_1^*)(n-9k_2^*) - (5n-9k_1^*-9k_2^*)^2}{16}$$

of the Hessian matrix of second-order partial derivatives of f. Finally, we take into account the maximum of f restricted to the boundary as well as the four corners $(1,1), (1, n-2), (n-2, 1), (n-2, n-2)$ of the domain defined by the inequalities $1 \leq k_1, k_2 \leq n-2$, and altogether, we discover that f attains its unique maximum at $(k_1, k_2) = \left(\frac{n}{3}, \frac{n}{3}\right)$. Consequently, we establish that

$$f(k_1, k_2) \leq f\left(\frac{n}{3}, \frac{n}{3}\right)$$

for all (k_1, k_2) such that $1 \leq k_1, k_2 \leq n-2$. Also, a direct calculation reveals that $f(n/3, n/3)$ is exactly $P(n)$. Thus, this completes the induction step.

Furthermore, the induction hypothesis tells us that $f(k_1, k_2) = P(n)$ if and only if for every non-leaf vertex v of T, the number of leaves in the three branches of the subtree of T rooted at v (the subtree induced by v together with all its descendants) are the same. In other words, $f(k_1, k_2) = P(n)$ if and only if each of the branches of T is a complete ternary tree, in which case T itself is a complete ternary tree. Hence, the limiting maximum density of C_3 in strictly ternary trees is

$$\lim_{n \to \infty} \frac{P(n)}{\binom{n}{3}} = \frac{3!}{24} = \frac{1}{4},$$

and this finishes the proof of the proposition. □

We point out that while the proof of Proposition 2.1.2.2 brings the statement that when $n = 3^h$ for some positive integer h, the maximum

$$\max_{\substack{|T|=n \\ T \text{ strictly ternary tree}}} c(C_3, T)$$

is realised for $T = CD_h^3$, it can be further demonstrated that the limit

$$\lim_{n \to \infty} \max_{\substack{|T|=n \\ T \text{ strictly ternary tree}}} \left(c(C_3, T) \binom{n}{3}^{-1} \right)$$

exists (thus, is $1/4$)—see the study conducted in the next chapter.

We conclude this section with a brief summary of some important computational rules on limits inferior and superior of sequences of real numbers. At present, we state them without proof—these can be found in many standard books of real analysis, for instance, [45; 46].

2.1.3 Algebraic properties of limits inferior and superior

A sequence $(\gamma_n)_{n\geq 1}$ of real numbers *converges* to a (finite) *limit* L (in which case we write $\lim_{n\to\infty} \gamma_n = L$) if its terms $\gamma_1, \gamma_2, \ldots$ eventually get and stay arbitrarily close to L for all sufficient large n. Formally, $\lim_{n\to\infty} \gamma_n = L$ if for every $\epsilon > 0$, there exists a natural number n_0 (allowed to depend on ϵ) such that $|\gamma_n - L| < \epsilon$ for all $n > n_0$. The limit of a convergent sequence is unique.

Every convergent sequence is bounded. Therefore, the boundedness is a necessary condition for a sequence to be convergent. Given a natural number N, the *tail* of the sequence $(\gamma_n)_{n\geq 1}$ is its (shifted) subsequence $\gamma_N, \gamma_{N+1}, \ldots$. The convergence properties of $(\gamma_n)_{n\geq 1}$ depend only on the behavior of $(\gamma_n)_{n\geq N}$ when we take N to be arbitrarily large. Thus, both sequences $(\gamma_n)_{n\geq 1}$ and $(\gamma_n)_{n\geq N}$ have the same limit for every choice of N. Consequently, taking away a finite number of terms in a sequence does not influence its convergence properties viz. boundedness and limit.

1. If γ_n and θ_n are two convergent sequences such that $\gamma_n \leq \theta_n$ for all sufficiently large n, then $\lim_{n\to\infty} \gamma_n \leq \lim_{n\to\infty} \theta_n$ (limits need not preserve strict inequalities). In particular, if γ_n is bounded from below by M_1 and above by M_2, we have $M_1 \leq \lim_{n\to\infty} \gamma_n \leq M_2$.

2. Bounded monotone sequences always converge: if γ_n is bounded and nondecreasing, then $\lim_{n\to\infty} \gamma_n = \sup_{n\geq 1} \gamma_n$; likewise, we have $\lim_{n\to\infty} \gamma_n = \inf_{n\geq 1} \gamma_n$ if γ_n is bounded and nonincreasing (note that both supremum and infimum exist and are finite for every bounded sequence[3]).

[3] This property is another way of expressing the completeness of real numbers in contrast to rational numbers, where the limit of a monotone bounded sequence of rational numbers need not be a rational number.

In order for the limit of a bounded (not necessarily monotone) sequence to always exist, one can introduce other concepts of limits: the *limit superior* denoted by lim sup and the *limit inferior* denoted by lim inf, which correspond to the limit of the supremum and the infimum of the tails of the sequence, respectively.

Definition 2.1.3.1. For a sequence $(\gamma_n)_{n\geq 1}$ of real numbers, the limit inferior and the limit superior are defined as follows:

$$\limsup_{n\to\infty} \gamma_n := \lim_{n\to\infty} \sup_{m\geq n} \gamma_m = \inf_n \sup_{m\geq n} \gamma_m \,,$$
$$\liminf_{n\to\infty} \gamma_n := \lim_{n\to\infty} \inf_{m\geq n} \gamma_m = \sup_n \inf_{m\geq n} \gamma_m \,.$$

The limit superior of γ_n is finite if $\sup\limits_{m\geq n} \gamma_m$ is finite for every n and bounded from below. Likewise, the limit inferior of γ_n is finite provided that $\inf\limits_{m\geq n} \gamma_m$ is finite for every n and bounded from above. Hence, both limits (inferior and superior) of every bounded sequence of real numbers exist and are finite.

Theorem 2.1.3.2. Let $(\gamma_n)_{n\geq 1}$ be a bounded real sequence. Then there is a subsequence of $(\gamma_n)_{n\geq 1}$ that converges to $\limsup\limits_{n\to\infty} \gamma_n$ ($\liminf\limits_{n\to\infty} \gamma_n$, respectively).

Proof. Set

$$L := \limsup_{n\to\infty} \gamma_n = \lim_{n\to\infty} \sup_{m\geq n} \gamma_m \,.$$

By the definition of limit, we know that for every $\epsilon > 0$, there exists a positive integer n_0 such that for every $n \geq n_0$, we have

$$\left| \left(\sup_{m\geq n} \gamma_m \right) - L \right| < \epsilon \,.$$

Since $L - \epsilon < \sup\limits_{m\geq l} \gamma_m$ for every $l \geq n_0$, there exists $k_l \geq l$ such that

$$L - \epsilon < \gamma_{k_l} \leq \sup_{m\geq l} \gamma_m < L + \epsilon$$

which implies that $|\gamma_{k_l} - L| < \epsilon$ for every $l \geq n_0$. But then, the subsequence $(\gamma_{k_l})_{k_l\geq l}$ converges to L. The statement on the limit inferior is proved along the same lines. $\square$

The following standard rules are relevant for us. We shall frequently refer back to them as needed, and this without further notice.

Proposition 2.1.3.3. For bounded sequences $(\gamma_n)_{n\geq 1}$ and $(\theta_n)_{n\geq 1}$ of real numbers, we have

1. $$\liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n \,.$$

   Moreover, the sequence $(\gamma_n)_{n\geq 1}$ is convergent and has the limit L if and only if

   $$\liminf_{n\to\infty} \gamma_n = \limsup_{n\to\infty} \gamma_n = L \,.$$

2. **Addition**:

   $$\begin{aligned}\limsup_{n\to\infty}(\gamma_n + \theta_n) &\leq \limsup_{n\to\infty} \gamma_n + \limsup_{n\to\infty} \theta_n \,,\\ \liminf_{n\to\infty}(\gamma_n + \theta_n) &\geq \liminf_{n\to\infty} \gamma_n + \liminf_{n\to\infty} \theta_n \,.\end{aligned}$$

   Furthermore, if θ_n converges to θ, then

   $$\limsup_{n\to\infty}(\gamma_n + \theta_n) = \limsup_{n\to\infty} \gamma_n + \theta \quad \text{and} \quad \liminf_{n\to\infty}(\gamma_n + \theta_n) = \liminf_{n\to\infty} \gamma_n + \theta \,.$$

3. **Multiplication**:

   $$\limsup_{n\to\infty}(\gamma_n \cdot \theta_n) \leq \Big(\limsup_{n\to\infty} \gamma_n\Big)\Big(\limsup_{n\to\infty} \theta_n\Big)$$

   if $(\gamma_n)_{n\geq 1}$ and $(\theta_n)_{n\geq 1}$ are both positive sequences. Moreover, if θ_n converges to θ, then

   $$\limsup_{n\to\infty}(\gamma_n \cdot \theta_n) = \theta \cdot \limsup_{n\to\infty} \gamma_n \,, \; \liminf_{n\to\infty}(\gamma_n \cdot \theta_n) = \theta \cdot \liminf_{n\to\infty} \gamma_n$$

   for $\theta \geq 0$, and

   $$\limsup_{n\to\infty}(\gamma_n \cdot \theta_n) = \theta \cdot \liminf_{n\to\infty} \gamma_n \,, \; \liminf_{n\to\infty}(\gamma_n \cdot \theta_n) = \theta \cdot \limsup_{n\to\infty} \gamma_n$$

   for $\theta \leq 0$.

We give a sketch of the proof of Proposition 2.1.3.3. According to Theorem 2.1.3.2, the sequence $(\gamma_n)_{n\geq 1}$ has at least one convergent subsequence, and moreover its has a subsequence that converges to $\limsup_{n\to\infty} \gamma_n$ (resp. $\liminf_{n\to\infty} \gamma_n$). Therefore, it suffices to let Γ denote the set of all the limits of the convergent subsequences of $(\gamma_n)_{n\geq 1}$ and prove that

$$\sup \Gamma \leq \limsup_{n\to\infty} \gamma_n, \text{ and } \inf \Gamma \geq \liminf_{n\to\infty} \gamma_n .$$

The conclusion is that

$$\max \Gamma = \limsup_{n\to\infty} \gamma_n, \text{ and } \min \Gamma = \liminf_{n\to\infty} \gamma_n$$

from which the computational rules enumerated in Proposition 2.1.3.3 follow.

2.2 Formal definition of the inducibility

In discrete mathematics and related areas of applications, one often wishes to find a specific submodel within a large structure. In this thesis, we focus on subtrees induced by leaves of a topological tree. By a *copy* of a topological tree S in another topological tree T, we mean any leaf-induced subtree of T isomorphic (in the sense of rooted tree isomorphism) to S. The total number of copies of S in T will be denoted by $c(S,T)$, and the quotient

$$\frac{c(S,T)}{\binom{|T|}{|S|}}$$

by $\gamma(S,T)$ whenever $|T| \geq |S|$. In plain words, $\gamma(S,T)$ is the proportion of all subsets of $|S|$ leaves of T that induce a copy of S. So $\gamma(S,T)$ lies between 0 and 1. For brevity, $\gamma(S,T)$ will be referred to as the *density* of S in T.

We are interested in the maximum of $\gamma(S,T)$ as the number of leaves of T grows to infinity, and the quantity

$$J(S) := \limsup_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T) = \limsup_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \gamma(S,T)\,,$$

where the maximum runs over all n-leaf topological trees, will be called the *inducibility of S in topological trees*.

When the maximum of the density $\gamma(D, T)$ of a d-ary tree D is taken over all d-ary trees T, we shall speak of the *inducibility in d-ary trees*. Specifically, the *inducibility* $I_d(D)$ of a d-ary tree D in d-ary trees is

$$I_d(D) = \limsup_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D,T) = \limsup_{\substack{|T|\to\infty \\ T \text{ } d\text{-ary tree}}} \gamma(D,T)\,.$$

Further, define the *inducibility of a d-ary tree D in strictly d-ary trees* as the limit superior, taken over all strictly d-ary trees, of the density of subsets of $|D|$ leaves that induce a copy of D. That is, we have

$$i_d(D) := \limsup_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D,T) = \limsup_{\substack{|T|\to\infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D,T)\,.$$

By observing that for $k \leq 2$, there is only one possibility for the k-leaf induced tree, we obtain the following simple fact:

Remark 2.2.0.1. The inducibility of the tree that has only one vertex as well as the 2-leaf topological tree is 1 in d-ary trees for every d.

We can interpret $J(S)$ as the *maximum asymptotic density* of copies of S in an arbitrary topological tree with sufficiently large number of leaves. Similarly, $I_d(D)$ (resp. $i_d(D)$) is the maximum asymptotic density of copies of D in an arbitrary d-ary tree (resp. strictly d-ary tree) with large enough number of leaves. Instead of determining the values of the maximum density

$$\max_{|T|=n} \gamma(S,T)$$

on every given number of leaves $n \geq |S|$ and among topological trees (or d-ary trees, or strictly d-ary trees), we shall only focus on understanding the asymptotic behaviour of such a density as n grows to infinity, which is therefore captured by the inducibility $J(S)$ (or $I_d(S)$, or $i_d(S)$ – depending on the underlying set over which the maximum is taken). Furthermore, we shall occasionally exhibit a sequence of trees that yields the inducibility in the limit.

Reasoning by means of induction will appear quite frequently as part of our proof techniques. In addition, we shall use the multinomial theorem and the so-called Muirhead's inequality in many places of the manuscript. Muirhead's inequality (also known as the bunching principle) is a useful tool employed to deal with seemingly difficult identities and inequalities of a certain sort. For the sake of completeness of the present exposition, we aim to discuss the fundamental idea behind the bunching principle.

2.3 The Bunching principle

Let $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_n)$ be vectors of nonnegative real numbers. Assume $a_1 \geq a_2 \geq \cdots \geq a_n$ and $b_1 \geq b_2 \geq \cdots \geq b_n$ in this order. We say that the vector A majorises the vector B if

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \,,$$

and for all $k = 1, 2, \ldots, n-1$,

$$\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} b_i \,.$$

Theorem 2.3.0.1 (**Muirhead's inequality**)**.** Consider a sequence $(x_1, x_2, \ldots, x_n)$ of positive real numbers. If $(a_1, a_2, \ldots, a_n)$ majorises $(b_1, b_2, \ldots, b_n)$ then it holds that

$$\sum_{\pi \in S_n} \prod_{i=1}^{n} x_{\pi(i)}^{a_i} \geq \sum_{\pi \in S_n} \prod_{i=1}^{n} x_{\pi(i)}^{b_i} \,,$$

where the sum is taken over the set S_n of all permutations of the indices $1, 2, \ldots, n$. There is equality if and only if either $a_i = b_i$ for all $i \in \{1, 2, \ldots, n\}$, or all the x_i are equal.

It is worth stressing that in its original form, "A majorises B" is a necessary but also sufficient condition in the theorem. A proof of this result can be found, for instance, in the book on inequalities by Hardy, Littlewood and Pólya [47]. Since it is a fundamental tool which we shall use to prove some of our main results, we find it important to give here a complete and comprehensive proof. The proof will follow after a series of very simple lemmas.

For $n \leq 2$, there is essentially nothing more to prove: the case $n = 1$ is trivial, while for the case $n = 2$, one can use the identity

$$\left(x_1^{a_1} \cdot x_2^{a_2} + x_2^{a_1} \cdot x_1^{a_2}\right) - \left(x_1^{b_1} \cdot x_2^{b_2} + x_2^{b_1} \cdot x_1^{b_2}\right) = \\ (x_1 \cdot x_2)^{a_2} \left(x_1^{b_1 - a_2} - x_2^{b_1 - a_2}\right)\left(x_1^{b_2 - a_2} - x_2^{b_2 - a_2}\right),$$

together with the fact that $b_1 \geq b_2 \geq a_2$.

Let $n \geq 3$ be fixed and assume the following conditions of the theorem are satisfied:

- (C1) $a_1 \geq a_2 \geq \cdots \geq a_n,\ b_1 \geq b_2 \geq \cdots \geq b_n$.
- (C2) $a_1 \geq a_2,\ a_1 + a_2 \geq b_1 + b_2,\ \ldots,\ a_1 + a_2 + \cdots + a_{n-1} \geq b_1 + b_2 + \cdots + b_{n-1}$.
- (C3) $a_1 + a_2 + \cdots + a_n = b_1 + b_2 + \cdots + b_n$.
- (C4) $A = (a_1, a_2, \ldots, a_n) \neq (b_1, b_2, \ldots, b_n) = B$.

Lemma 2.3.0.2. Let $n \geq 3$. The following hold:

1. (L1) There exists at least two different indices i such that $a_i \neq b_i$.
2. (L2) At least one index i satisfying $a_i > b_i$ exists.
3. (L3) At least one index i satisfying $a_i < b_i$ exists.
4. (L4) The first index i such that $a_i \neq b_i$ yields $a_i > b_i$.
5. (L5) There are indices j, k such that

$$a_j > b_j,\ a_{j+1} = b_{j+1},\ \ldots,\ a_{k-1} = b_{k-1},\ a_k < b_k\,.$$

6. (L6) A vector $A' := (a'_1, a'_2, \ldots, a'_n)$ of real numbers satisfying both

$$\left|\{i \in \{1, 2, \ldots, n\} : a'_i \neq b_i\}\right| < \left|\{i \in \{1, 2, \ldots, n\} : a_i \neq b_i\}\right|$$

and

$$\sum_{\pi \in S_n} \prod_{i=1}^{n} x_{\pi(i)}^{a'_i} \leq \sum_{\pi \in S_n} \prod_{i=1}^{n} x_{\pi(i)}^{a_i}$$

can be constructed from the vectors A and B so that, if further vectors A'', A''' etc. are constructed in the same way, B is obtained after a finite number of steps.

Proof. The proof goes as follows:

1. Condition (C4) guarantees the existence of at least one index i, with $a_i \neq b_i$. At least two such indices exist as (C3) is to be satisfied.

2. If $a_i \leq b_i$ for all i, then

$$a_1 \leq a_2,\ a_1 + a_2 \leq b_1 + b_2,\ \ldots,\ a_1 + a_2 + \cdots + a_n \leq b_1 + b_2 + \cdots + b_n\,,$$

and together with (C2), this implies that

$$a_1 = b_1,\ a_1 + a_2 = b_1 + b_2,\ \ldots,\ a_1 + a_2 + \cdots + a_n = b_1 + b_2 + \cdots + b_n\,,$$

contradicting (C4).

3. From (L2), let j be such that $a_j > b_j$. If $a_i \geq b_i$ for all $i \neq j$, then

$$a_1 + \cdots + a_{j-1} + a_{j+1} + \cdots + a_n \geq b_1 + \cdots + b_{j-1} + b_{j+1} + \cdots + b_n\,,$$

implying that $a_1 + \cdots + a_n > b_1 + \cdots + b_n$, which contradicts (C3).

4. If the first i, $1 \leq i \leq n$, such that $a_i \neq b_i$ gave rise to $a_i < b_i$, then we would have $a_1 = b_1, \ldots, a_{i-1} = b_{i-1}$ and so $a_1 + \cdots + a_i < b_1 + \cdots + b_i$, contradicting (C2).

5. From (L3), select the first index k satisfying $a_k < b_k$ and from (L4), select the greatest index j less than k with $a_j > b_j$.

6. Let j, k be indices as defined in (L5). We are going to operate exclusively on the indices j and k. To this end, we define a new vector A' by setting $a'_i := a_i$ for $i \neq \{j, k\}$. Thus, we have

$$\sum_{\pi \in S_n} \left(\prod_{i=1}^{n} x_{\pi(i)}^{a_i} - \prod_{i=1}^{n} x_{\pi(i)}^{a'_i} \right) = \frac{1}{2} \sum_{\pi \in S_n} \Bigg[\Bigg(\prod_{\substack{i=1 \\ i \neq \{j,k\}}}^{n} x_{\pi(i)}^{a_i} \Bigg) \cdot \left(x_{\pi(j)}^{a_j} \cdot x_{\pi(k)}^{a_k} + x_{\pi(k)}^{a_j} \cdot x_{\pi(j)}^{a_k} - x_{\pi(j)}^{a'_j} \cdot x_{\pi(k)}^{a'_k} - x_{\pi(k)}^{a'_j} \cdot x_{\pi(j)}^{a'_k} \right) \Bigg]\,.$$

We can choose a'_j and a'_k in such a way that $a'_j + a'_k = a_j + a_k$ so that we obtain a factor of the form $\left(x^{\alpha}_{\pi(j)} - x^{\alpha}_{\pi(k)}\right)$: so we have

$$\sum_{\pi\in S_n}\left(\prod_{i=1}^{n} x^{a_i}_{\pi(i)} - \prod_{i=1}^{n} x^{a'_i}_{\pi(i)}\right) =$$
$$\frac{1}{2}\cdot\sum_{\pi\in S_n}\Big(\prod_{\substack{i=1\\ i\neq\{j,k\}}}^{n} x^{a_i}_{\pi(i)}\Big)\left(x^{a'_j-a_k}_{\pi(j)} - x^{a'_j-a_k}_{\pi(k)}\right)x^{a_k}_{\pi(j)}\left(x^{a'_k-a_k}_{\pi(j)} - x^{a'_k-a_k}_{\pi(k)}\right)x^{a_k}_{\pi(k)}\,,$$

which is readily seen to be nonnegative. Now we need to choose a'_j and a'_k in such a way that the condition

$$\left|\{i\in\{1,2,\ldots,n\} : a'_i \neq b_i\}\right| < \left|\{i\in\{1,2,\ldots,n\} : a_i \neq b_i\}\right|$$

is met, i.e., guaranteeing either $a'_j = b_j$ or $a'_k = b_k$. Since $j < k$, we have $a_k \leq a_j$ and $b_k \leq b_j$ (so $a_k < b_k \leq b_j < a_j$). Therefore, $a'_j - a'_k = 2b_j - (a_j + a_k)$ if $a'_j = b_j$, and $a'_j - a'_k = (a_j + a_k) - 2b_k$ if $a'_k = b_k$. Altogether, we can take

$$a'_j - a'_k = \begin{cases} 2b_j - (a_j + a_k) & \text{if } |2b_j - (a_j + a_k)| \geq |a_j + a_k - 2b_k|\,, \\ a_j + a_k - 2b_k & \text{otherwise}\,. \end{cases}$$

This finishes the proof of the lemma.

□

Proof of Theorem 2.3.0.1. Replace A with A' in the proof of Lemma 2.3.0.2 and call $A'' := A^{(2)}$ the vector obtained from $A' := A^{(1)}$ and B, and so on. Since the number of indices i such that $a'_i \neq b_i$ diminishes (by at least 1) each time we apply the transformation which gives A' using elements of A and B, we shall get to a stage (after a finite number of steps, say l) where $a^{(l)}_i = b_i$ for all $1 \leq i \leq n$, and so the process comes to an end: the first part of the theorem is proved. In view of the relation

$$\sum_{\pi\in S_n}\left(\prod_{i=1}^{n} x^{a_i}_{\pi(i)} - \prod_{i=1}^{n} x^{a'_i}_{\pi(i)}\right) = \frac{1}{2}\sum_{\pi\in S_n}\Bigg[\Big(\prod_{\substack{i=1\\ i\neq\{j,k\}}}^{n} x^{a_i}_{\pi(i)}\Big)$$
$$\cdot\left(x^{a'_j-a_k}_{\pi(j)} - x^{a'_j-a_k}_{\pi(k)}\right)x^{a_k}_{\pi(j)}\left(x^{a'_k-a_k}_{\pi(j)} - x^{a'_k-a_k}_{\pi(k)}\right)x^{a_k}_{\pi(k)}\Bigg] \geq 0\,,$$

it is clearly seen that for $A \neq B$, equality is secured in the theorem if and only if $x_{\pi(j)} = x_{\pi(k)}$ for every $\pi \in S_n$, i.e., $x_1 = x_2 = \cdots = x_n$. □

Example 2.3.0.3. Since the vector $(5,1,0,0)$ majorises the vector $(2,2,1,1)$, we get

$$\sum_{\pi\in S_4} x_{\pi(1)}^5 \cdot x_{\pi(2)}^1 \cdot x_{\pi(3)}^0 \cdot x_{\pi(4)}^0 \geq \sum_{\pi\in S_4} x_{\pi(1)}^2 \cdot x_{\pi(2)}^2 \cdot x_{\pi(3)}^1 \cdot x_{\pi(4)}^1$$

for every sequence (x_1, x_2, x_3, x_4) of positive real numbers, i.e.,

$$\begin{aligned} &x_1^5 \cdot x_2 + x_1^5 \cdot x_3 + x_1^5 \cdot x_4 + x_1 \cdot x_2^5 + x_2^5 \cdot x_3 + x_2^5 \cdot x_4 + x_1 \cdot x_3^5 + x_2 \cdot x_3^5 + x_3^5 \cdot x_4 \\ &\quad + x_1 \cdot x_4^5 + x_2 \cdot x_4^5 + x_3 \cdot x_4^5 \geq\ 2(x_1^2 \cdot x_2^2 \cdot x_3 \cdot x_4 + x_1^2 \cdot x_2 \cdot x_3^2 \cdot x_4 \\ &\qquad + x_1^2 \cdot x_2 \cdot x_3 \cdot x_4^2 + x_1 \cdot x_2^2 \cdot x_3^2 \cdot x_4 + x_1 \cdot x_2^2 \cdot x_3 \cdot x_4^2 + x_1 \cdot x_2 \cdot x_3^2 \cdot x_4^2) \end{aligned}$$

with equality if and only if $x_1 = x_2 = x_3 = x_4$.

Some known inequalities can be deduced from the inequality of Muirhead. For instance, the well-known inequality between the geometric mean and the arithmetic mean can be inferred from it:

Example 2.3.0.4 (Geometric-Arithmetic Mean Inequality)**.** Let $(x_1, x_2, \ldots, x_n)$ be a sequence of $n \geq 1$ positive real numbers. Then the inequality

$$\begin{aligned} \frac{x_1 + x_2 + \cdots + x_n}{n} &= \mathrm{AM}(x_1,x_2,\ldots,x_n) \\ &\geq \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \mathrm{GM}(x_1, x_2, \ldots, x_n) \end{aligned}$$

holds with equality if and only if $x_1 = x_2 = \cdots = x_n$.

Proof. Since $(1,\underbrace{0,0,\ldots,0}_{(n-1)\ 0's})$ majorises $\big(\underbrace{1/n, 1/n, \ldots, 1/n}_{n \text{ terms}}\big)$, we have

$$(n-1)! \cdot (x_1 + x_2 + \cdots + x_n) \geq n! \cdot x_1^{\frac{1}{n}} \cdot x_2^{\frac{1}{n}} \cdot \ldots \cdot x_n^{\frac{1}{n}}$$

by virtue of Muirhead's inequality. □

We mention that there is no simple recipe that leads to a solution when computing the inducibility. Quite often, we shall not have an explicit formula for the maximum number of copies $\max_{|T|=n} c(S,T)$ of S among n-leaf trees, even for S in the family of trees of interest.

Chapter 3

Inducibility of d-ary trees

In this chapter, we study the inducibility in d-ary trees (rooted trees whose vertex outdegrees are bounded from above by $d \geq 2$). We determine the exact inducibility for stars and so-called binary caterpillars. For T in the family of strictly d-ary trees (every vertex has 0 or d children), we prove that the difference between the maximum density $\gamma(D, T)$ of a d-ary tree D in T and the inducibility $i_d(D)$ of D is of order at most $\mathcal{O}(|T|^{-1/2})$ compared to the more general case $I_d(D)$ where it is shown that the difference is $\mathcal{O}(|T|^{-1})$ which, in particular, responds positively to an existing conjecture from [1] on the inducibility in binary trees. We also discover that the inducibility of a binary tree (2-ary tree) in d-ary trees is independent of d. Furthermore, we establish a general lower bound on the inducibility of every d-ary tree and also provide a bound for some special trees. Moreover, we find that the maximum inducibility is attained only for binary caterpillars (binary trees for which removal of all the leaves yields a path) for every d.

The material is based on the results from the following paper [34]: *Inducibility of d-ary trees. É. Czabarka, A. A. V. Dossou-Olory, L. A. Székely and S. Wagner. Preprint* `https://arxiv.org/abs/1802.03817`

3.1 Introduction

We recall that if S is a subset of the leaf set of a d-ary tree T, then the unique subtree obtained by first extracting the minimal subtree of T containing the leaves in S, and then suppressing, except possibly for the root, the vertices of degree 2 in the induced tree, is called a leaf-induced subtree of T; see Figure 3.1. Moreover, the resulting tree has a root in a natural way, namely the most recent common ancestor of the leaves in S.

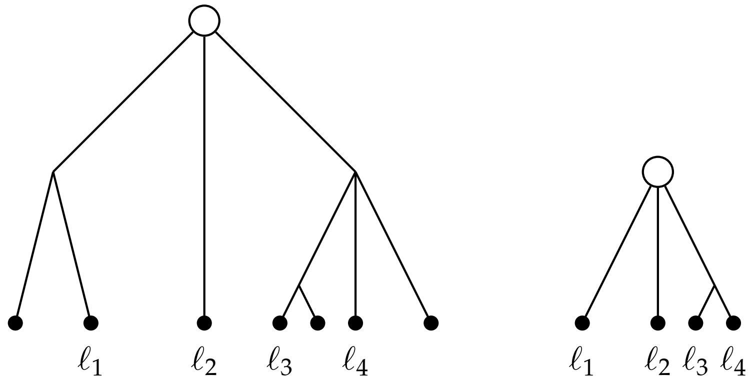


Figure 3.1: A ternary tree and the subtree induced by the four leaves $\{\ell_1, \ell_2, \ell_3, \ell_4\}$.

A copy of D in T is any leaf-induced subtree of T isomorphic (in the sense of rooted trees) to D. The total number of copies of D in T is denoted by $c(D,T)$, and the quotient $c(D,T)/\binom{|T|}{|D|}$ by $\gamma(D,T)$, which is the proportion of all subsets of $|D|$ leaves of T that induce a copy of D. The inducibility of a d-ary tree D in d-ary trees T is

$$I_d(D) = \limsup_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(D,T)\,, \tag{3.1.1}$$

where we mean

$$\limsup_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(D,T) = \limsup_{n\to\infty} \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T).$$

This concept of inducibility is, of course, in analogy to the very first one [2] for simple graphs. However, it is not clear at this point whether the sequence

$$\Big(\max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T) \Big)_{n \geq |D|}$$

converges for every d-ary tree D, in which case we could simply write

$$I_d(D) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T)\,.$$

But notice that equation (3.1.1) already tells us that the maximum number of copies of D in an arbitrary n-leaf d-ary tree is at most

$$\binom{n}{|D|}\big(I_d(D) + o(1)\big)$$

as $n \to \infty$. In fact, we shall prove later that $I_d(D)$ is always a positive real number for every d-ary tree D—see Proposition 3.5.0.1. For the particular case where $d = 2$, it was conjectured in [1] that the asymptotic formula

$$\max_{\substack{|T|=n \\ T\ \text{binary tree}}} \gamma(B,T) = I_2(B) + \mathcal{O}(n^{-1}) \tag{3.1.2}$$

holds for every binary tree B. In the present chapter, we affirm this conjecture, and even generalise the result for every d (see Theorem 3.3.0.1).

It turns out that $I_d(D)$ can also be computed by merely taking the limit superior in the family of strictly d-ary trees. Recall that the inducibility of a d-ary tree D in strictly d-ary trees is

$$i_d(D) = \limsup_{\substack{|T|\to\infty \\ T\ \text{strictly}\ d\text{-ary tree}}} \gamma(D,T)\,.$$

We note that $0 \leq i_d(D) \leq I_d(D) \leq 1$ by definition. We demonstrate in Section 3.3 that the identity $I_d(D) = i_d(D)$ holds for every d-ary tree D and every d (Theorem 3.3.0.3). As counterpart of equation (3.1.2), we show in Corollary 3.3.0.4 that

$$\max_{\substack{|T|=n \\ T\ \text{strictly}\ d\text{-ary tree}}} \gamma(D,T) = i_d(D) + \mathcal{O}(n^{-1/2}).$$

In Section 3.4, we prove that the inducibility of any fixed binary tree is the same among binary trees and d-ary trees, for any d; in Section 3.5 we show that the inducibility of any d-ary tree is positive, and finally in Section 3.6 we prove that the only d-ary trees that have inducibility 1 are the binary caterpillars.

3.2 On stars and binary caterpillars

As mentioned in the introduction, it is instructive to treat the single vertex (of the tree that has only one vertex) as both leaf and root. By joining leaves to a single vertex v, one obtains a *star* with root v. We denote the k-leaf star by C_k as suggested by Figure 3.2.

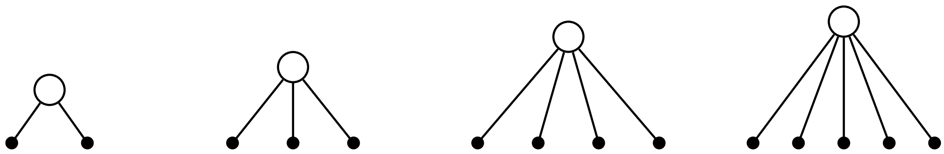


Figure 3.2: Stars C_k.

It is common knowledge that in several instances, stars, so-called caterpillars, and paths are solutions to many extremal problems among trees with respect to a given combinatorial invariant. For instance, the star minimises the Hosoya index (number of independent edge subsets) and thus maximises the Merrifield-Simmons index (number of independent vertex subsets) among connected graphs all having the same number of vertices [48; 49; 50]. We begin our investigations on the inducibility with these families of trees (the latter is not a d-ary tree!).

We recall that a complete d-ary tree is a strictly d-ary tree in which all the leaves (d^h in total) are at the same distance h from the root. For our next theorem, which characterises the inducibility of stars, we need Muirhead's inequality (already discussed in Chapter 2): this inequality states that

$$\sum_{\pi \in S_d} \prod_{j=1}^{d} x_{\pi(j)}^{l_j} \geq \sum_{\pi \in S_d} \prod_{j=1}^{d} x_{\pi(j)}^{m_j}$$

for all nonnegative real numbers $x_1, x_2, \ldots, x_d$ whenever the vector $(l_1, l_2, \ldots, l_d)$ majorises the vector $(m_1, m_2, \ldots, m_d)$. The sums on both sides are taken over all permutations of the indices $1, 2, \ldots, d$.

Theorem 3.2.0.1. For every fixed positive integer $d \geq 2$ and every $k \in \{2, 3, \ldots, d\}$, the inducibility of the k-leaf star C_k in strictly d-ary trees is

$$i_d(C_k) = \frac{d!}{(d-k)!(d^k - d)}.$$

Moreover, $i_d(C_k)$ is an increasing function in d for every $k \geq 3$, starting with $d = k$.

Proof. Fix $d \geq 2$, and let us prove that for a strictly d-ary tree T with n leaves, we have

$$c(C_k, T) \leq \binom{d}{k} \frac{n^k - n}{d^k - d}$$

for every $k \geq 2$, with equality if T is a complete d-ary tree.

For $k = 2$, the assertion is immediate as $c(C_2, T) = \binom{|T|}{2}$. For $k \geq 3$, we use induction on n. The base cases $n < k$ are trivial as there cannot be any copies of C_k in T. For the induction step, consider the d branches $T_1, T_2, \ldots, T_d$ of T with $n_1, n_2, \ldots, n_d$ leaves, respectively. So we have $n_1 + n_2 + \cdots + n_d = n$. We distinguish possible scenarios that can occur for a subset of k leaves:

- all k leaves belong to the same branch of T. The total number of these subsets of leaves that induce C_k is given by

$$c(C_k, T_1) + c(C_k, T_2) + \cdots + c(C_k, T_d)\,,$$

- one of the branches of T contains more than one of the k leaves, but not all of them. In this case the leaf-induced subtree is not isomorphic to C_k,

- k of the branches of T, say (without loss of generality) $T_1, T_2, \ldots, T_k$, contain exactly one of the leaves each. In this case, the k leaves always induce C_k, yielding $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ copies of C_k.

Therefore, we establish that

$$c(C_k, T) = \sum_{i=1}^{d} c(C_k, T_i) \; + \sum_{\{i_1, i_2, \ldots, i_k\} \subseteq \{1,2,\ldots,d\}} \prod_{j=1}^{k} n_{i_j}\,.$$

The induction hypothesis gives

$$c(C_k, T) \leq \frac{\binom{d}{k}}{d^k - d} \sum_{i=1}^{d} (n_i^k - n_i) \; + \sum_{\{i_1, i_2, \ldots, i_k\} \subseteq \{1,2,\ldots,d\}} \prod_{j=1}^{k} n_{i_j}\,. \tag{3.2.1}$$

On the other hand, the multinomial theorem yields the following decomposition:

$$\begin{aligned}
n^k = \left(\sum_{i=1}^{d} n_i\right)^k &= \sum_{\substack{l_1,l_2,\dots,l_d \geq 0 \\ l_1+l_2+\cdots+l_d=k}} \binom{k}{l_1,l_2,\dots,l_d} \prod_{i=1}^{d} n_i^{l_i} \\
&= \sum_{\substack{l_1,l_2,\dots,l_d \geq 0 \\ l_1+l_2+\cdots+l_d=k}} \binom{k}{l_1,l_2,\dots,l_d} \frac{1}{d!} \sum_{\pi \in S_d} \prod_{i=1}^{d} n_{\pi(i)}^{l_i} \\
&= \sum_{i=1}^{d} n_i^k + \sum_{\substack{0 \leq l_1,l_2,\dots,l_d < k \\ l_1+l_2+\cdots+l_d=k}} \binom{k}{l_1,l_2,\dots,l_d} \frac{1}{d!} \sum_{\pi \in S_d} \prod_{i=1}^{d} n_{\pi(i)}^{l_i} .
\end{aligned}$$

Since every vector $(l_1, l_2, \dots, l_d)$ of nonnegative integers with $l_1 + l_2 + \cdots + l_d = k$ majorises the vector $(1, 1, \dots, 1, 0, 0, \dots, 0)$ (k ones, followed by $d - k$ zeros), we can apply Muirhead's inequality to every term in the second sum of this decomposition:

$$\sum_{\pi \in S_d} \prod_{i=1}^{d} n_{\pi(i)}^{l_i} \geq \sum_{\pi \in S_d} \prod_{i=1}^{k} n_{\pi(i)} = k! \cdot (d-k)! \sum_{\{i_1,i_2,\dots,i_k\} \subseteq \{1,2,\dots,d\}} \prod_{j=1}^{k} n_{i_j} .$$

This gives us

$$\begin{aligned}
n^k &\geq \sum_{i=1}^{d} n_i^k + \sum_{\substack{0 \leq l_1,l_2,\dots,l_d < k \\ l_1+l_2+\cdots+l_d=k}} \binom{k}{l_1,l_2,\dots,l_d} \frac{k! \cdot (d-k)!}{d!} \sum_{\{i_1,i_2,\dots,i_k\} \subseteq \{1,2,\dots,d\}} \prod_{j=1}^{k} n_{i_j} \\
&= \sum_{i=1}^{d} n_i^k + (d^k - d) \frac{k! \cdot (d-k)!}{d!} \sum_{\{i_1,i_2,\dots,i_k\} \subseteq \{1,2,\dots,d\}} \prod_{j=1}^{k} n_{i_j} ,
\end{aligned}$$

using the multinomial theorem in the opposite direction now. We can rewrite this as

$$\sum_{\{i_1,i_2,\dots,i_k\} \subseteq \{1,2,\dots,d\}} \prod_{j=1}^{k} n_{i_j} \leq \frac{\binom{d}{k}}{d^k - d} \left(n^k - \sum_{i=1}^{d} n_i^k\right) .$$

Plugging this into (3.2.1) yields

$$\begin{aligned}
c(C_k, T) &\leq \frac{\binom{d}{k}}{d^k - d} \sum_{i=1}^{d} (n_i^k - n_i) + \frac{\binom{d}{k}}{d^k - d} \left(n^k - \sum_{i=1}^{d} n_i^k\right) \\
&= \frac{\binom{d}{k}}{d^k - d} (n^k - n) ,
\end{aligned}$$

completing the induction. Furthermore, equality can only arise in this context if $n_1 = n_2 = \cdots = n_d$. Thus the inequality holds with equality if and only if, for every internal vertex v of T, the number of leaves in the d branches of the subtree of T rooted at v are the same: in this case, T is a complete d-ary tree as well. The assertion on the inducibility follows by passing to the density and taking the limit:

$$i_d(C_k) = \lim_{n\to\infty} \frac{\binom{d}{k} \cdot \left(n^k - n\right)/\left(d^k - d\right)}{\binom{n}{k}} = \frac{d!}{(d-k)!(d^k - d)}\,.$$

We now turn to the second assertion of the theorem.

Claim: For every given positive integer $k \geq 3$,

$$u_k(x) = \frac{(x-1)\cdot(x-2)\cdot\ldots\cdot(x-k+1)}{x^{k-1}-1}$$

is an increasing function for $x \geq k$.

Proof of the Claim: Indeed, we have

$$\log u_k(x) = -\log(x^{k-1}-1) + \sum_{i=1}^{k-1} \log(x-i)$$

so that

$$\begin{aligned}
\frac{d}{dx}\Big(\log u_k(x)\Big) &= -\frac{k-1}{x^{k-1}-1}\cdot x^{k-2} + \sum_{i=1}^{k-1}\frac{1}{x-i} \\
&= \sum_{i=1}^{k-1}\Big(\frac{1}{x-i} - \frac{x^{k-2}}{x^{k-1}-1}\Big) \\
&= \frac{1}{x^{k-1}-1}\sum_{i=1}^{k-1}\Big(\frac{-1+i\cdot x^{k-2}}{x-i}\Big) > 0
\end{aligned}$$

as $x \geq k \geq 3$. We conclude that $\big(i_d(C_k)\big)_{d\geq k}$ is an increasing sequence as soon as $k \geq 3$. □

We remark that the special case $d = k = 3$ in Theorem 3.2.0.1 coincides with Proposition 2.1.2.2 which was proved by other means in Chapter 2.

The next family of d-ary trees of interest is the family of *binary caterpillars*. We determine the exact inducibility of every binary caterpillar, which in turn, is also demonstrated to be the same among d-ary trees for every d.

By a *d-ary caterpillar*, we mean a strictly d-ary tree in which every non-leaf vertex has $d-1$ children that are leaves, except for the lowest, which has d children that are leaves. Note that the non-leaf vertices must lie on a single path starting at the root. We denote the d-ary caterpillar with k leaves by F_k^d (refer to Figure 3.3 for ternary caterpillars).

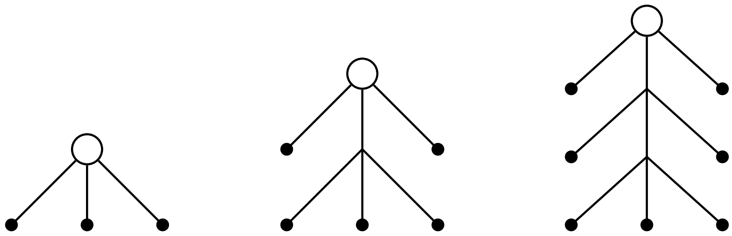


Figure 3.3: Ternary caterpillars F_k^3.

Theorem 3.2.0.2. Let $d \geq 2$ be an arbitrary but fixed positive integer. For the binary caterpillar F_k^2, we have

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(F_k^2, T) = 1 - \mathcal{O}(n^{-1})$$

for every k, d and all $n \geq k$. In particular, $i_d(F_k^2) = I_d(F_k^2) = 1$ for every k and d.

Proof. Fix $d \geq 2$. We use a direct counting argument which gives us a lower bound on the number of copies of the binary caterpillar F_k^2 in the d-ary caterpillar F_n^d. A binary caterpillar can be constructed by attaching exactly one pendant edge to all vertices of a path, except the last one (that is furthest away from the root). Thus, a copy of F_k^2 in F_n^d can be obtained by first choosing a subset of k vertices from the set of internal vertices of F_n^d, and then for each of them, choosing one of its (at least) $d-1$ children that are leaves. Since every strictly d-ary tree T has exactly $(|T|-1)/(d-1)$ internal vertices [1], we deduce that there are at least

$$\binom{\frac{n-1}{d-1}}{k}(d-1)^k$$

[1]It is proved by induction on $|T|$.

copies of F_k^2 in F_n^d. Therefore, we get

$$c(F_k^2, F_n^d) \geq \binom{\frac{n-1}{d-1}}{k}(d-1)^k = \frac{n^k}{k!} - \mathcal{O}(n^{k-1})\,.$$

Hence, because we have

$$c(F_k^2, F_n^d) \leq \binom{n}{k} \leq \frac{n^k}{k!}$$

by definition, the assertion on the maximum density of F_k^2 in strictly d-ary trees and thus the inducibility $i_d(F_k^2)$ follows. In particular, we obtain

$$i_d(F_k^2) = I_d(F_k^2) = 1\,.$$

□

It can actually be shown that

$$c(F_k^2, F_n^d) = (d-1)^{k-1}\binom{\frac{n-1}{d-1}}{k-1} \cdot \frac{2n-(d-2)(k-2)}{2k} \tag{3.2.2}$$

for $k, n > 1$, but this precision becomes immaterial when computing the inducibility $i_d(F_k^2)$. The proof of this identity can be found at the end of this chapter (Section 3.7).

3.3 Inducibility vs. maximum density

Our first goal in this section is to prove a conjecture from [1], which states that the maximum of $\gamma(D,T)$ over trees T with n leaves converges fairly quickly to the inducibility $I_d(D)$, namely at a rate of n^{-1}. (To be precise, the conjecture was only made in the binary case.)

Theorem 3.3.0.1. For every fixed positive integer $d \geq 2$ and every d-ary tree D, we have

$$\max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T) = I_d(D) + \mathcal{O}(n^{-1})$$

for all $n \geq |D|$. In particular,

$$\lim_{n\to\infty} \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T) = I_d(D)\,,$$

where the limit is that of a decreasing sequence.

Proof. Fix $d \geq 2$. Let D and T be arbitrary d-ary trees such that $|D| \leq |T|$. Since $c(D, T)$ represents the number of subsets of $|D|$ leaves of T that induce a copy of D, we immediately deduce that the $|T|$ leaves of T are contained in $|D| \cdot c(D, T)/|T|$ copies of D on average. Thus, there exist leaves l_1 and l_2 of T satisfying the relation

$$c_{l_1}(D, T) \leq \frac{|D| \cdot c(D, T)}{|T|} \leq c_{l_2}(D, T) \tag{3.3.1}$$

where $c_l(D, T)$ stands for the number of l-containing subsets of $|D|$ leaves of T that induce a copy of D.

- We can assume $|T| \geq 2$. The number of subsets of $|D|$ leaves of T not involving the leaf l_1 that induce a tree isomorphic to D is exactly $c(D, T) - c_{l_1}(D, T)$, which is therefore at least $\left(1 - \frac{|D|}{|T|}\right) \cdot c(D, T)$ in view of relation (3.3.1). Create from T a new tree T' by suppressing the leaf l_1 (and suppress its former neighbor, if its new degree is 2). It follows that

  $$\max_{|T'|=n-1} c(D, T') \geq \left(1 - \frac{|D|}{n}\right) \max_{|T|=n} c(D, T)$$

  as T was taken to be an arbitrary d-ary tree. Hence, by passing to the density by dividing by $\binom{n-1}{|D|}$, we obtain

  $$\max_{\substack{|T'|=n-1 \\ T' \ d\text{-ary tree}}} \gamma(D, T') \geq \max_{\substack{|T|=n \\ T \ d\text{-ary tree}}} \gamma(D, T)$$

  for every $n \geq 1 + |D|$. So the sequence

  $$\Big(\max_{\substack{|T|=n \\ T \ d\text{-ary tree}}} \gamma(D, T) \Big)_{n \geq |D|}$$

  is decreasing and bounded from below, which means that the limit

  $$\lim_{n \to \infty} \max_{\substack{|T|=n \\ T \ d\text{-ary tree}}} \gamma(D, T)$$

  exists and is $I_d(D)$.

- Denote by T^+ the d-ary tree obtained by replacing the leaf l_2 by an internal vertex with two leaves l_2, l_2' attached to it. That way, the number of copies of D in T^+ not involving l_2' is just $c(D,T)$, whereas the number of copies of D in T^+ involving l_2' is no less than the number of copies of D in T involving l_2. Therefore, by relation (3.3.1), the quantity $\left(1+\frac{|D|}{|T|}\right)\cdot c(D,T)$ offers a natural lower bound on $c(D,T^+)$. It follows that

$$\left(1+\frac{|D|}{n}\right)\max_{|T|=n} c(D,T) \leq \max_{|T|=n+1} c(D,T)$$

as T was assumed to be an arbitrary n-leaf d-ary tree. Consequently, passing to the density, we obtain

$$\max_{\substack{|T|=n+1\\ T\ d\text{-ary tree}}} \gamma(D,T) \geq \left(1-\frac{|D|(-1+|D|)}{n(n+1)}\right) \max_{\substack{|T|=n\\ T\ d\text{-ary tree}}} \gamma(D,T)$$

for every $n \geq |D|$, and by p-fold iteration

$$\max_{\substack{|T|=n+p\\ T\ d\text{-ary tree}}} \gamma(D,T) \geq$$

$$\max_{\substack{|T|=n\\ T\ d\text{-ary tree}}} \gamma(D,T)\cdot\prod_{j=0}^{p-1}\left(1-\frac{|D|(-1+|D|)}{(n+p-j)(n+p-j-1)}\right)$$

for all n,p with $p \geq 1$ and $n \geq |D|$. Fixing $n \geq |D|$ and $p \geq 1$, we have

$$0 \leq \frac{|D|(-1+|D|)}{(n+p-j)(n+p-j-1)} < 1$$

for every $0 \leq j \leq p-1$. A simple induction on p yields

$$\prod_{j=0}^{p-1}\left(1-\frac{|D|(-1+|D|)}{(n+p-j)(n+p-j-1)}\right) \geq$$

$$1-\sum_{j=0}^{p-1}\frac{|D|(-1+|D|)}{(n+p-j)(n+p-j-1)}.$$

Now, letting $p \to \infty$ instantly gives the estimate

$$I_d(D) \geq \max_{\substack{|T|=n\\ T\ d\text{-ary tree}}} \gamma(D,T)\cdot\left(1-\sum_{i=0}^{\infty}\frac{|D|(-1+|D|)}{(n+i+1)(n+i)}\right)$$

for every $n \geq |D|$. Since

$$\sum_{i=0}^{\infty} \frac{1}{(n+i+1)(n+i)} = \sum_{i=0}^{\infty} \left(\frac{1}{n+i} - \frac{1}{n+i+1}\right) = \frac{1}{n},$$

this shows that

$$I_d(D) \geq \left(1 - \frac{|D|(-1+|D|)}{n}\right) \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T).$$

Now we combine the two contributions to obtain

$$0 \leq \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T) - I_d(D) \leq \frac{|D|(-1+|D|)}{n}$$

for every $n \geq |D|$. The desired asymptotic formula follows immediately. $\square$

We remark that the averaging reasoning employed in the proof of Theorem 3.3.0.1 does not work for a strictly d-ary tree. For example, removing one leaf (as well as the single edge incident to it) from a strictly d-ary tree yields a tree that is no longer strictly d-ary as soon as $d \geq 3$. Moreover, the error term $\mathcal{O}(n^{-1})$ is generally best possible, as e.g. the discussion of the complete binary tree of height 2 (with four leaves) in [1], or the star with three leaves in Proposition 2.1.2.2 in Chapter 2 shows.

Our next focus is to prove that $i_d(D)$ and $I_d(D)$ are always equal for every d-ary tree D and every d.

Lemma 3.3.0.2. For a fixed positive integer k, we have

$$\frac{\binom{n}{k}}{\binom{p+n}{k}} = 1 - \mathcal{O}(p/n)$$

as $p \geq 1$ and $n/p \to \infty$.

Proof. We have

$$\begin{aligned}
\frac{\binom{n}{k}}{\binom{p+n}{k}} &= \frac{n(n-1)\dots(n-k+1)}{(n+p)(n+p-1)\dots(n+p-k+1)} \\
&\geq \frac{(n-k)^k}{(n+p)^k} = \left(1 - \frac{k}{n}\right)^k \left(1 + \frac{p}{n}\right)^{-k} \\
&\geq \left(1 - \frac{k^2}{n}\right)\left(1 - \frac{k \cdot p}{n}\right)
\end{aligned}$$

where the last inequality follows from Bernoulli's inequality. This implies that

$$\begin{aligned}\frac{\binom{n}{k}}{\binom{p+n}{k}} &\geq \left(1-\frac{k^2}{n}\right)\left(1-\frac{k\cdot p}{n}\right) \geq 1-\frac{k\cdot p+k^2}{n}\\ &\geq 1-\frac{2k^2\cdot p}{n}\end{aligned}$$

as it is easy to see that $p+k \leq 2p\cdot k$. This completes the proof of the lemma. □

Theorem 3.3.0.3. Fix a positive integer $d \geq 2$, and let D be a d-ary tree. For every positive integer $n \equiv 1 \mod (d-1)$ and every d-ary tree T with $\lfloor n^{\frac{1}{2}} \rfloor$ leaves, there exists a strictly d-ary tree T^* with n leaves such that the asymptotic formula

$$\gamma(D,T) = \gamma(D,T^*) + \mathcal{O}(n^{-1/2})$$

holds as $n \to \infty$, and the $\mathcal{O}$-constant depends on d only. In particular, we have

$$i_d(D) = I_d(D)\,.$$

Proof. Fix $d \geq 2$ and $n \equiv 1 \mod (d-1)$. Consider an arbitrary d-ary tree T (not necessarily a strictly d-ary tree) such that $|T| = \lfloor n^{\frac{1}{2}} \rfloor = n^{\frac{1}{2}} - \mathcal{O}(1)$. We describe an explicit construction for T^*; the line of reasoning follows probabilistic ideas.

- For every $r \in \{2,3,\ldots,d-1\}$, add $d-r$ more branches of one leaf each to every internal vertex of T whose number of children is r. Call the augmented tree T', and denote by $\tilde{L}(T')$ the set of the additional leaves added to T to obtain the strictly d-ary tree T'. If $|T|_r$ stands for the number of internal vertices of T with r children, then we have

$$|\tilde{L}(T')| = \sum_{r=2}^{d-1}(d-r)|T|_r < (d-2)|T|,\ |\tilde{L}(T')| = \mathcal{O}\big(n^{\frac{1}{2}}\big)$$

because it is easy to see (by induction on $|T|$) that the total number of internal vertices of T is less than its number of leaves. Note that

$$|T'| = |T| + |\tilde{L}(T')| = \Theta\big(n^{\frac{1}{2}}\big)\,.$$

- Consider the tree T'. Let m be the greatest positive integer satisfying both $m \equiv 1 \mod (d-1)$ and $m \leq n^{\frac{1}{2}}$. So we have $\lfloor n^{\frac{1}{2}} \rfloor - (d-2) \leq m \leq \lfloor n^{\frac{1}{2}} \rfloor$. Since it suffices to prove the statement for sufficiently large n, we may assume that $m \geq d$. Choose an arbitrary strictly d-ary tree S with $m-(d-1)$ leaves so that

$$\lfloor n^{\frac{1}{2}} \rfloor + 3 - 2 \cdot d \leq |S| \leq \lfloor n^{\frac{1}{2}} \rfloor + 1 - d\,.$$

Append a copy of S to every leaf l of T' that does not belong to $\tilde{L}(T')$ by identifying its root with l. Call the resulting tree T''. We shall refer to the tree S as a 'dangling' tree of T''. Note that

$$|T''| = |T| \cdot |S| + |\tilde{L}(T')| = n - \mathcal{O}\big(n^{\frac{1}{2}}\big)$$

and that $|T''| < n$ in view of the inequalities

$$|T| \leq n^{\frac{1}{2}}, \ |S| \leq n^{\frac{1}{2}} - (d-1), \ |\tilde{L}(T')| < (d-1)|T|\,.$$

- Additionally, pick an arbitrary strictly d-ary tree S_P with

$$1 + n - |T''| = \mathcal{O}\big(n^{\frac{1}{2}}\big)$$

leaves and append the root of T'' to a leaf of S_P. Denote by T^* the strictly d-ary tree that results from this construction. Note that

$$|T^*| = |T''| + |S_P| - 1 = n\,.$$

A picture that shows this construction is given in Figure 3.4.

Now let $1 \leq k \leq |T|$ be an arbitrary but fixed positive integer, and pick k leaves of T^* uniformly at random. The probability that none of the k randomly chosen leaves of T^* lies in $\tilde{L}(T')$ or S_P and no two of them belong to the same dangling tree of T^* is exactly

$$\frac{\binom{|T|}{k}}{\binom{|T^*|}{k}} \cdot |S|^k\,.$$

In words: since there are exactly $|T|$ dangling trees in T^*, we choose k of them and one leaf from each of the k chosen dangling trees to obtain such

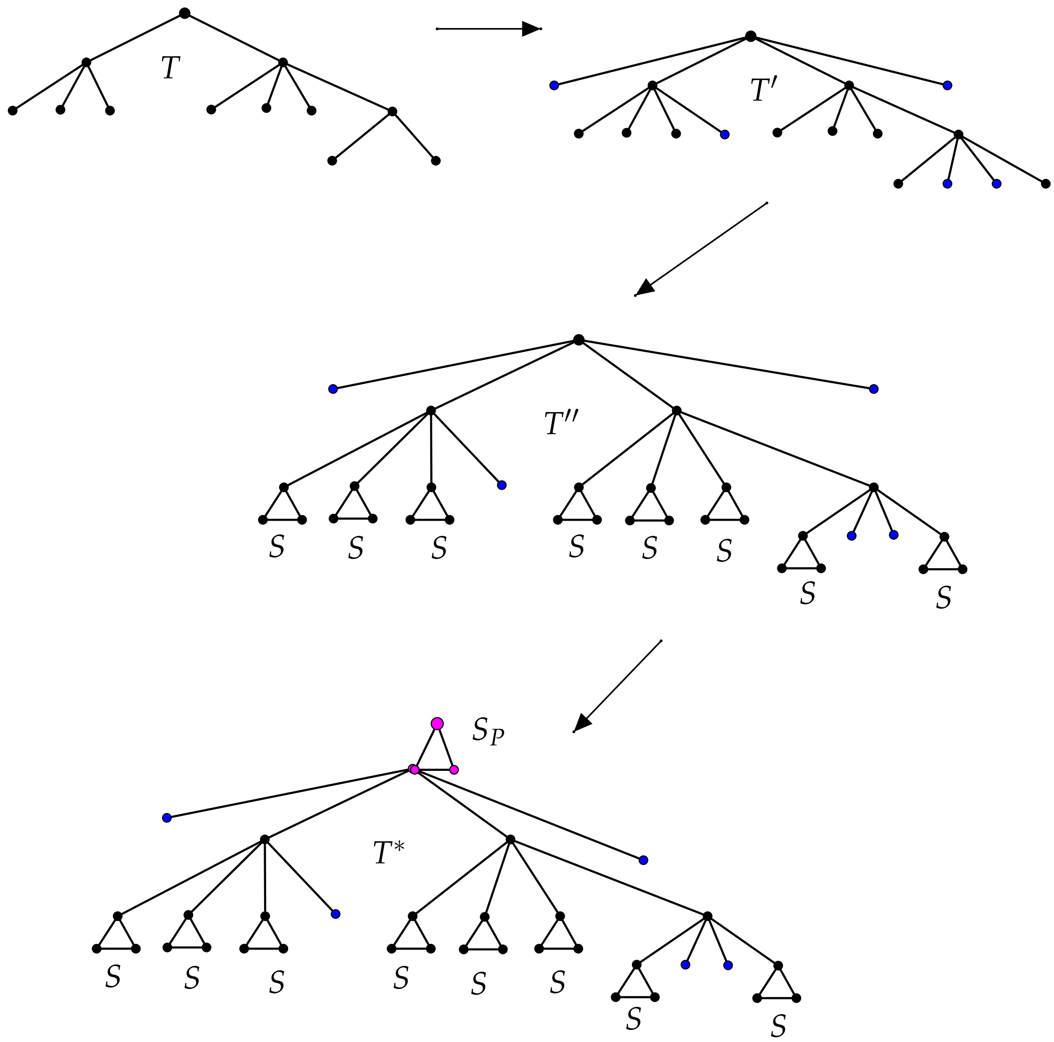


Figure 3.4: A step-by-step illustration of the explicit construction given in the proof of Theorem 3.3.0.3: T^* is a strictly 4-ary tree obtained from a 4-ary tree T with eight leaves.

a subset of k leaves of T^*. With Lemma 3.3.0.2 at our disposal, we obtain

$$\begin{aligned}\frac{\binom{|T|}{k}}{\binom{|T^*|}{k}} \cdot |S|^k &= \frac{\big(|T| \cdot |S|\big)^k \big(1 - \mathcal{O}(|T|^{-1})\big)}{n^k \big(1 - \mathcal{O}(n^{-1})\big)} \\ &= 1 - \mathcal{O}\big(n^{-\frac{1}{2}}\big)\end{aligned} \tag{3.3.2}$$

as $n \to \infty$. Now let D be a d-ary tree with k leaves. Note that the tree induced by k leaves of T^* that belong to k distinct dangling trees is equal to the tree induced by the k leaves of T to which these k dangling tree were attached. Hence the probability that k randomly chosen leaves of T^* are from distinct dangling trees of T^* and induce a copy of D is given by

$$\frac{c(D,T)}{\binom{n}{k}} \cdot |S|^k$$

(recall that T^* has n leaves). From this observation and the fact that $\gamma(D,T^*)$ is exactly the probability that k distinct randomly chosen leaves of T^* induce a copy of D, we deduce that

$$\gamma(D,T^*) = \frac{c(D,T)}{\binom{n}{k}} \cdot |S|^k + Q\Big(1 - \frac{\binom{|T|}{k} \cdot |S|^k}{\binom{n}{k}}\Big)$$

by virtue of the law of total probability, where Q stands for the probability that k distinct leaves of T^* induce a copy of D under the condition that the event 'k randomly chosen leaves of T^* are from distinct dangling trees of T^*' has not occurred. This implies that

$$\gamma(D,T^*) = \frac{c(D,T)}{\binom{|T|}{k}} + \Big(1 - \frac{\binom{|T|}{k} \cdot |S|^k}{\binom{n}{k}}\Big)\Big(Q - \frac{c(D,T)}{\binom{|T|}{k}}\Big),$$

and since Q and $\gamma(D,T) = c(D,T)/\binom{|T|}{k}$ are both between 0 and 1, it follows from the asymptotic formula (3.3.2) that

$$\gamma(D,T^*) - \gamma(D,T) = \mathcal{O}(n^{-1/2}).$$

This finishes the proof of the first part of the theorem. Finally, the immediate consequence we obtain is that

$$I_d(D) = \limsup_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(D,T) = \limsup_{|T^*|\to\infty} \gamma(D,T^*) \leq \limsup_{\substack{|T|\to\infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D,T) = i_d(D).$$

In other words, this shows that $I_d(D) \leq i_d(D)$. Thus, the proof of the second part of the theorem is completed as well because we have $I_d(D) \geq i_d(D)$ by definition. □

With Theorem 3.3.0.3 and its proof at hand, we can now prove an analogue of Theorem 3.3.0.1 for the maximum density in strictly d-ary trees.

Corollary 3.3.0.4. For every fixed positive integer $d \geq 2$ and every d-ary tree D, we have

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D,T) = i_d(D) + \mathcal{O}(n^{-1/2})$$

for every $n \geq |D|$. In particular, we have

$$\lim_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D,T) = i_d(D)\,.$$

Proof. Using the identity $i_d(D) = I_d(D)$ from Theorem 3.3.0.3 together with the first claim of Theorem 3.3.0.1, namely

$$\max_{\substack{|T|=n \\ T \ d\text{-ary tree}}} \gamma(D,T) \leq I_d(D) + \mathcal{O}\big(n^{-1}\big)\,,$$

we immediately get

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D,T) \leq i_d(D) + \mathcal{O}\big(n^{-1}\big)\,. \tag{3.3.3}$$

Also, the second claim of Theorem 3.3.0.1 gives

$$i_d(D) = I_d(D) \leq \max_{\substack{|T|=\lfloor n^{1/2}\rfloor \\ T \ d\text{-ary tree}}} \gamma(D,T)\,, \tag{3.3.4}$$

while the first claim of Theorem 3.3.0.3 guarantees a particular strictly d-ary tree T^* on n leaves for the maximiser tree in (3.3.4), such that

$$\gamma(D,T) = \gamma(D,T^*) + \mathcal{O}\big(n^{-\frac{1}{2}}\big)\,. \tag{3.3.5}$$

Formulas (3.3.4) and (3.3.5) immediately give

$$i_d(D) \leq \max_{\substack{|T'|=n \\ T' \text{ strictly } d\text{-ary tree}}} \gamma(D,T') + \mathcal{O}\big(n^{-\frac{1}{2}}\big)\,,$$

which, together with (3.3.3), completes the proof of the corollary. □

In certain cases, we suspect a stronger asymptotic result on the maximum density in strictly d-ary trees—see Chapter 5. This, in particular, happens when

$$\max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} c(D,T)$$

is attained by strictly d-ary trees for every $n \equiv 1 \mod (d-1)$. Finally, a natural question to ask at this point is the following:

QUESTION 3.3.0.5. Can the $\mathcal{O}$-term in Corollary 3.3.0.4 be improved somewhat further for general d-ary trees D?

3.4 Inducibility of a binary tree in d-ary trees

Our aim in this section is to compare the inducibilities of a binary tree B in d-ary trees, for different values of d. It turns out that $i_d(B)$ is independent of d.

Theorem 3.4.0.1. Every binary tree B satisfies

$$i_d(B) = i_2(B) = I_d(B) = I_2(B)$$

for every d.

Proof. The following correspondence is our stepping stone for proving the assertion. Let $d \geq 3$ be fixed, and fix an arbitrary total order $\prec$ on the set of all strictly d-ary trees. For a strictly d-ary tree T, we always order its branches $T_1, T_2, \ldots, T_d$ in such a way that

$$T_1 \preceq T_2 \preceq \cdots \preceq T_d\,.$$

From the tree T, we build a binary tree $G(T)$ by means of the recursive algorithm depicted in Figure 3.5, starting with the single leaf being invariant under G. More specifically, for $|T| > 1$, the tree $G(T)$ is obtained as follows:

- draw a path on $d-1$ vertices $v_2, v_3, \ldots, v_d$ in this order (v_2 and v_d are the endvertices of the path);

- attach a leaf l_i (by dropping a pendant edge) to every vertex v_i of the path except for the lowest vertex v_2;
- replace the vertex v_2 of the path by an internal vertex with two leaves l_1 and l_2 attached to it;
- identify the root of $G(T_i)$ to leaf l_i for every $i \in \{1, 2, \ldots, d\}$; the vertex v_d is the root of $G(T)$.

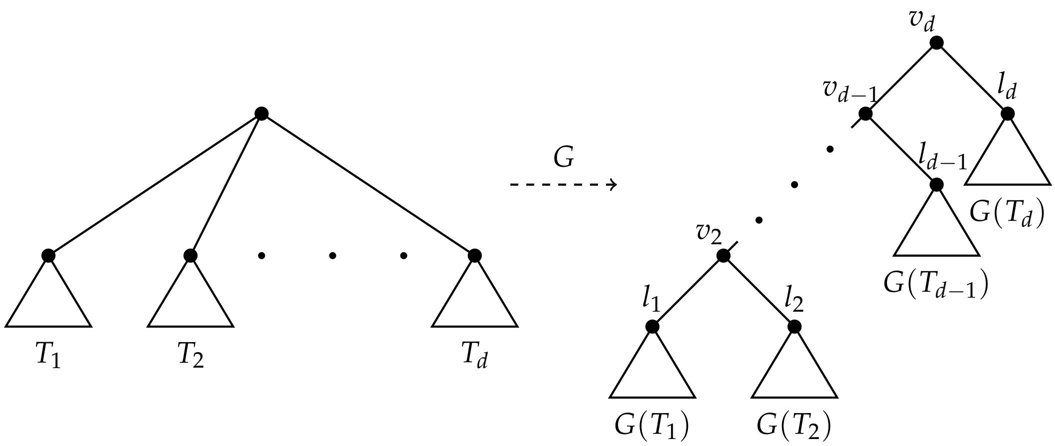


Figure 3.5: T is a strictly d-ary tree and $G(T)$ is its corresponding binary tree under the tree-transformation G described in the proof of Theorem 3.4.0.1.

For example, G maps the star C_d to the binary caterpillar F_d^2. With this transformation at hand, let us prove that the inequality

$$c\big(B, G(T)\big) \geq c(B, T)$$

holds for every binary tree B and every strictly d-ary tree T.

The construction of $G(T)$ also yields a natural bijection between the leaves of T and those of $G(T)$. We can show by induction that if a set of leaves of T induces a copy of a binary tree B, then so do the corresponding leaves in $G(T)$. This is clearly true when $|T| = 1$. For the induction step, we have two cases:

- If the leaves that induce B all lie in one branch T_i for some $i \in \{1, 2, \ldots, d\}$, then the corresponding leaves lie in $G(T_i)$, and we are done by the induction hypothesis.

- Otherwise, they lie in exactly two branches T_i and T_j for some $i, j \in \{1, 2, \ldots, d\}$ such that $i \neq j$, and the two leaf sets induce the two branches of B. By the induction hypothesis, this is also true for the corresponding leaves in $G(T_i)$ and $G(T_j)$, and we are done again.

This shows that every subset of leaves of T that induces a copy of B corresponds to a unique subset of leaves of $G(T)$ that induces a copy of B. Therefore, there is an injection from the copies of B in T to the copies of B in $G(T)$, and we have $c(B, T) \leq c\big(B, G(T)\big)$ as claimed. Consequently, we arrive at

$$i_d(B) \leq \limsup_{\substack{|T| \to \infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma\big(B, G(T)\big) \leq \limsup_{\substack{|T'| \to \infty \\ T' \text{ binary tree}}} \gamma\big(B, T'\big) = i_2(B)\,.$$

On the other hand, we have both $i_2(B) = I_2(B)$ and $I_d(B) \geq I_2(B)$ by definition, while Corollary 3.3.0.4 gives us $i_d(B) = I_d(B)$. Hence, the statement of the theorem follows. $\square$

We point out that the analogue of Theorem 3.4.0.1 for non-binary trees is not true in general: as Theorem 3.2.0.1 shows, the inducibility $i_d(C_k)$ of the k-leaf star is a strictly increasing function of d for $k \geq 3$.

From here onwards, we shall use only $i_d(D)$. Our next section deals with some bounds on the inducibility.

3.5 Some general results

Let us say something about how small $i_d(D)$ can be:

Proposition 3.5.0.1. Let $d \geq 2$ be an arbitrary but fixed positive integer. Every d-ary tree D with at least two leaves satisfies

$$i_d(D) \geq \frac{(-1 + |D|)!}{-1 + |D|^{-1+|D|}}\,.$$

In particular, every d-ary tree has positive inducibility.

Proof. We employ a tree-construction similar to the one used for proving that $i_d(D) = I_d(D)$. Fix $d \geq 2$. For two strictly d-ary trees S_1, S_2, denote by $\mathcal{F}(S_1; S_2)$ the unique strictly d-ary tree that is formed by appending the root of S_2 to every leaf of S_1; see Figure 3.6 for a picture. So we have $|\mathcal{F}(S_1; S_2)| = |S_1| \cdot |S_2|$ by definition.

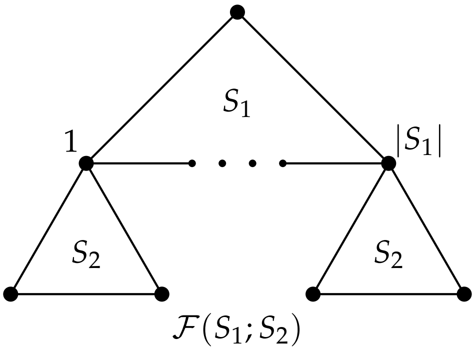


Figure 3.6: A rough picture of the tree $\mathcal{F}(S_1; S_2)$ defined in the proof of Proposition 3.5.0.1.

Starting with $T^{[1]}$ being the single leaf, we define recursively the family $T^{[n+1]} = \mathcal{F}(D; T^{[n]})$ of strictly d-ary trees. It is clear that $T^{[2]} = D$ and $|T^{[n]}| = |D|^{-1+n}$. In this fashion, there are $|D| \cdot c(D, T^{[n]})$ copies of D in the $|D|$ copies of $T^{[n]}$ that are attached to the leaves of D to obtain $T^{[n+1]}$. Also, if one takes one arbitrary leaf from each of these $|D|$ copies, one obtains a total of another $|T^{[n]}|^{|D|}$ copies of D. Thus, we get

$$c(D, T^{[1+n]}) \geq |D| \cdot c(D, T^{[n]}) + |T^{[n]}|^{|D|},$$

which gives

$$c(D, T^{[1+n]}) \geq \sum_{j=0}^{n-1} |D|^j \cdot |T^{[-j+n]}|^{|D|}$$

by iteration as $c(D, T^{[1]}) = 0$. Since $|T^{[-j+n]}| = |D|^{-1-j+n}$, we establish the inequality

$$c(D, T^{[n]}) \geq |D|^{(-2+n) \cdot |D|} \sum_{j=0}^{n-2} |D|^{j(1-|D|)},$$

which becomes

$$\gamma(D, T^{[n]}) \geq \frac{|D|^{(-1+n)|D|} - |D|^{-1+n}}{(|D|^{|D|} - |D|)\binom{|D|^{-1+n}}{|D|}}.$$

Letting $n \to \infty$ produces

$$\limsup_{n\to\infty} \gamma\big(D, T^{[n]}\big) \geq \frac{(-1+|D|)!}{-1+|D|^{-1+|D|}}$$

and, in particular, the assertion of the proposition. □

The bound obtained in Proposition 3.5.0.1 seems to be weak in general. Possibly, C_d is even the only d-ary tree that attains this bound (see Theorem 3.2.0.1).

The tree $\mathcal{F}(S_1; S_2)$ (as constructed in the proof of Proposition 3.5.0.1) offers another special result:

Theorem 3.5.0.2. Let $d \geq 2$ be an arbitrary but fixed positive integer. For two d-ary trees S_1, S_2, let $\mathcal{F}(S_1; S_2)$ be as defined in the proof of Proposition 3.5.0.1. Then we have

$$i_d\big(\mathcal{F}(S_1; S_2)\big) \geq \frac{(|S_1| \cdot |S_2|)!}{(|S_2|!)^{|S_1|} \cdot |S_1|^{|S_1| \cdot |S_2|}} \big(i_d(S_2)\big)^{|S_1|}.$$

Proof. Fix $d \geq 2$. For a d-ary tree T, it is easy to see that a copy of $\mathcal{F}(S_1; S_2)$ in $\mathcal{F}(S_1; T)$ can be obtained by taking one copy of S_2 in T for each of the $|S_1|$ copies of T planted in $\mathcal{F}(S_1; T)$. Thus, the inequality

$$c\big(\mathcal{F}(S_1; S_2), \mathcal{F}(S_1; T)\big) \geq c\big(S_2, T\big)^{|S_1|}$$

is valid so that we obtain

$$\max_{\substack{|T'|=|T| \\ T' \ d\text{-ary tree}}} \frac{c\big(\mathcal{F}(S_1; S_2), \mathcal{F}(S_1; T')\big)}{\binom{|S_1| \cdot |T'|}{|S_1| \cdot |S_2|}} \geq \frac{\binom{|T|}{|S_2|}^{|S_1|}}{\binom{|S_1| \cdot |T|}{|S_1| \cdot |S_2|}} \cdot \gamma\big(S_2, T\big)^{|S_1|}.$$

Now we choose T in such a way that

$$|T| = n, \ c(S_2, T) = \max_{\substack{|T''|=n \\ T'' \ d\text{-ary tree}}} c\big(S_2, T''\big).$$

Letting $n \to \infty$ yields

$$\limsup_{n\to\infty} \max_{\substack{|T'|=n \\ T' \ d\text{-ary tree}}} \gamma\big(\mathcal{F}(S_1; S_2), \mathcal{F}(S_1; T')\big) \geq$$

$$\frac{(|S_1| \cdot |S_2|)!}{(|S_2|!)^{|S_1|} \cdot |S_1|^{|S_1| \cdot |S_2|}} \big(i_d(S_2)\big)^{|S_1|}.$$

Consequently, we obtain the desired inequality. □

In some special cases, there is an upper bound that asymptotically matches the lower bound on $i_d\big(\mathcal{F}(S_1;S_2)\big)$ as $|\mathcal{F}(S_1;S_2)|$ gets large—see Chapter 5.

3.6 The d-ary trees with the maximal inducibility

The first natural question to pose regarding a graph invariant concerns its extreme values and the extremal graphs. In our case, we have already proved that every d-ary tree has positive inducibility (Proposition 3.5.0.1) in d-ary trees and that $i_d(F_k^2) = 1$ for every k and every d (Theorem 3.2.0.2). But besides binary caterpillars, are there other d-ary trees with inducibility 1? The answer turns out to be negative. This extends the result for binary trees in [1].

Theorem 3.6.0.1. Let $d \geq 2$ be an arbitrary but fixed positive integer. Among d-ary trees, only binary caterpillars have inducibility 1.

Proof. Fix $d \geq 2$ and let us prove that for any fixed positive integers $k > 1$ and $n > d^{k-2}$, every d-ary tree with n leaves contains a copy of the k-leaf binary caterpillar F_k^2.

The key observation is that if a d-ary tree has height at least $k-1$, then it must have F_k^2 as a leaf-induced subtree. So fix $k > 1$, $n > d^{k-2}$, and consider a d-ary tree T with n leaves. It is easy to see by induction on h that for every fixed integer $h \geq 0$, a d-ary tree with height at most h can never have more than d^h leaves. Since T has more than d^{k-2} leaves, its height must be at least $k-1$, so it contains F_k^2. In conclusion, for $k > 1$, every n-leaf d-ary tree must contain a copy of the binary caterpillar F_k^2 as soon as $n > d^{k-2}$.

Now, consider a d-ary tree D that has at least three leaves. If D is different from $F_{|D|}^2$, then we obtain $c(D,T) < \binom{|T|}{|D|}$ for every d-ary tree T with n leaves provided that $n > d^{|D|-2}$. Therefore, fixing $n > d^{|D|-2}$, we get

$$\max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T) < 1\,.$$

On the other hand, from the second part of Theorem 3.3.0.1, we have

$$i_d(D) \leq \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T)\,,$$

which now completes the proof of the theorem. □

As one would expect, the limit

$$\liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(F_k^2, T)$$

is positive for every d and every k. An exact formula is available in Chapter 7.

3.7 Proof of identity 3.2.2 and concluding comments

We want to prove that for every choice of $d \geq 2$, the identity

$$c(F_k^2, F_n^d) = (d-1)^{k-1} \binom{\frac{n-1}{d-1}}{k-1} \frac{2n-(d-2)(k-2)}{2k}$$

holds for all $k, n > 1$.

For $k = 2$, this identity gives $c(F_2^2, F_n^d) = n(n-1)/2$ which is true as $c(F_2^2, F_n^d) = \binom{n}{2}$. For $n = d$, we obtain $c(F_k^2, F_d^d) = 0$ if $k > 2$ and $c(F_k^2, F_d^d) = d(d-1)/2$ if $k = 2$, which is also true. So we can continue by induction on n. To achieve this, we use the recurrence relation

$$c(F_k^2, F_n^d) = c(F_k^2, F_{n-d+1}^d) + (d-1)c(F_{k-1}^2, F_{n-d+1}^d)$$

which is obtained by noticing that only two scenarios can occur for a subset of k leaves of the n-leaf d-ary caterpillar F_n^d:

- either all the k leaves of F_k^2 belong to the branch F_{n-d+1}^d of the tree F_n^d giving $c(F_k^2, F_{n-d+1}^d)$ copies of F_k^2,
- or all the $k-1$ leaves of the branch F_{k-1}^2 of the tree F_k^2 lie in the leaf-set of F_{n-d+1}^d while the single leaf branch of F_k^2 is taken from any of the $d-1$ branches of F_n^d which are leaves: this yields $(d-1) \cdot c(F_{k-1}^2, F_{n-d+1}^d)$ copies of F_k^2.

The induction hypothesis gives us

$$\begin{aligned} c(F_k^2, F_n^d) &= (d-1)^{k-1} \binom{\frac{n-d}{d-1}}{k-1} \frac{2(n-d+1)-(d-2)(k-2)}{2k} \\ &\quad + (d-1)^{k-1} \binom{\frac{n-d}{d-1}}{k-2} \frac{2(n-d+1)-(d-2)(k-3)}{2(k-1)}. \end{aligned}$$

Using the identities

$$\binom{\frac{n-d}{d-1}}{k-1} = \Big(1 - \frac{(k-1)(d-1)}{n-1}\Big) \binom{\frac{n-1}{d-1}}{k-1}$$

and

$$\binom{\frac{n-d}{d-1}}{k-2} = \frac{(k-1)(d-1)}{n-1} \binom{\frac{n-1}{d-1}}{k-1},$$

we obtain

$$\begin{aligned} c(F_k^2, F_n^d) &= \frac{(d-1)^{k-1}}{2k} \binom{\frac{n-1}{d-1}}{k-1} \Big(2(n-d+1)-(d-2)(k-2) \\ &\quad - \frac{(k-1)(d-1)}{n-1}\big(2(n-d+1)-(d-2)(k-2)\big) \\ &\quad + \frac{k(d-1)}{n-1}\big(2(n-d+1)-(d-2)(k-3)\big)\Big) \\ &= \frac{(d-1)^{k-1}}{2k} \binom{\frac{n-1}{d-1}}{k-1} \Big(2n-(d-2)(k-2) + \frac{2n(d-1)}{n-1} \\ &\quad - 2(d-1)\Big(1+\frac{d-1}{n-1}\Big) + \frac{2(d-1)(d-2)}{n-1}\Big) \\ &= (d-1)^{k-1} \binom{\frac{n-1}{d-1}}{k-1} \frac{2n-(d-2)(k-2)}{2k} \end{aligned}$$

completing the induction step.

It seems appropriate to close this chapter with further problems on the inducibility of d-ary trees.

Since we already have a general lower bound on $i_d(D)$, it would be nice to understand the following problem:

PROBLEM 3.7.0.1. Given positive integers $d \geq 2$ and $k \geq 5$, find

$$\min_{\substack{|D|=k \\ D\ d\text{-ary tree}}} i_d(D),$$

and furthermore, characterise the d-ary trees that attain this minimum.

We conjecture that except for binary trees, the inducibility in d-ary trees always depends on d (in contrast to binary trees, where it does not; see Theorem 3.4.0.1 and the discussion thereafter).

CONJECTURE 3.7.0.2. Let $d \geq 2$ be an arbitrary but fixed positive integer. Among all d-ary trees D, only binary trees satisfy

$$i_d(D) = i_{d+1}(D) = i_{d+2}(D) = \cdots .$$

Chapter 4

Inducibility of topological trees

Trees without vertices of degree 2 are sometimes named topological trees. In this chapter, we investigate the inducibility of (rooted) topological trees with k leaves. The inducibility of a topological tree S is the limit superior of the proportion of all subsets of $|S|$ leaves of T that induce a copy of S as the size of T grows to infinity. In particular, this relaxes the degree-restriction for the notion of the inducibility in d-ary trees presented in Chapter 3. We discuss some of the properties of this generalised concept and investigate its connection with the degree-restricted inducibility. In addition, we prove that stars and binary caterpillars are the only topological trees that have an inducibility of 1. We also find an explicit lower bound on the limit inferior of the proportion of all subsets of k leaves of T that induce either a star or a binary caterpillar with k leaves as the size of T tends to infinity.

The material is based on the following paper [35]: *Inducibility of topological trees. A. A. V. Dossou-Olory and S. Wagner. Accepted for publication in Quaestiones Mathematicae*.

4.1 Introduction and auxiliary results

In Chapter 3, we proposed an extension of the inducibility of binary trees to d-ary trees for every $d \geq 2$. The object of this chapter is to continue the investigation of the inducibility of trees. In particular, it is natural to consider a variant of the concept that relaxes the degree-restriction on the vertices of the tree: one might be interested in knowing how large the number of appearances of a tree in another larger tree with a given number of leaves can be.

Before we can present our results, we need to recall some formal definitions from Chapter 2. We call any rooted tree in which every internal vertex has outdegree at least 2 a topological tree. For two topological trees S, T, we denote by $c(S, T)$ the number of copies of S in T, which is the number of subsets of the leaf set of T that induce a tree isomorphic to S in the sense of rooted trees (i.e., the isomorphism maps the root of one tree to the root of the other tree). For $|T| \geq |S|$, the quotient $\gamma(S, T) = c(S, T)/\binom{|T|}{|S|}$ is referred to as the density of S in T.

We are interested in the maximum of $\gamma(S, T)$ in the limit: the quantity

$$J(S) = \limsup_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S, T) = \limsup_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \gamma(S, T), \tag{4.1.1}$$

where the maximum runs over all topological trees, is called the inducibility of S in topological trees.

A topological tree with the property that every vertex has outdegree no more than d (≥ 2) is called a d-ary tree. We simply call a 2-ary tree a binary tree. When the maximum of the density $\gamma(D, T)$ of a d-ary tree D in T is taken over all d-ary trees, we speak of the inducibility in d-ary trees which is given by

$$I_d(D) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T),$$

as presented in Chapter 3, where the limit is shown to exist. We shall prove that this is also the case for $J(S)$, see Theorem 4.2.0.1. The d-ary trees that attain the maximum inducibility 1 are determined in Chapter 3: for every $d \geq 2$, the maximal trees are binary caterpillars (paths with one pendant edge dropped from all the vertices except for one of the endvertices). Things change, however, when the vertices of the tree are allowed to have any outdegree, see Theorem 4.2.0.2.

Clearly, for any given d-ary tree D, the inducibility $I_d(D)$ is non-decreasing with respect to d, and thus tends to a definite limit as $d \to \infty$. In the following section, we establish that $J(S) = \lim_{d\to\infty} I_d(S)$ for every topological tree S. We also prove a link between the maximum density of S in topological trees and the actual inducibility $J(S)$, and demonstrate that stars and binary caterpillars are the only topological trees with the maximum inducibility 1. Furthermore, we find an explicit lower bound on the limit inferior of the proportion of all subsets of k leaves of T that induce either a star or a binary caterpillar with k leaves as the size of the topological tree T grows to infinity (see Proposition 4.2.0.4). In addition, we show that one can obtain a lower bound on the inducibility of any topological tree.

4.2 Main results and discussion

As for the degree-restricted inducibility, we begin our investigation by providing an estimate on how much the general inducibility can differ from the maximum density $\gamma(S,T)$ in topological trees. The result is an analogue of Theorem 3.3.0.1 in Chapter 3 for $J(S)$. It asserts that for every topological tree S, the maximum of $\gamma(S,T)$ tends to a definite limit as $|T| \to \infty$ and that the precise gap between the maximum density and the limit is of order at most $\mathcal{O}(|T|^{-1})$.

Theorem 4.2.0.1. Let S be a topological tree. Then the double inequality

$$0 \leq \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T) - J(S) \leq |S|(-1+|S|)n^{-1}$$

is valid for all $n \geq |S|$. In particular, we have

$$J(S) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)\,.$$

Proof. We can follow the same averaging argument used in Chapter 3 to prove Theorem 3.3.0.1. Since we shall make heavy use of the intermediary results that appear in the proof, we present them for completeness.

Let S and T be two topological trees such that $|S| \leq |T|$. We write $L(T)$ for the set of leaves of T. For $l \in L(T)$, let us denote by $c_l(S,T)$ the number of l-containing subsets of leaves of T that induce a copy of S. Thus, since

$$\sum_{l \in L(T)} c_l(S,T) = |S| \cdot c(S,T)\,,$$

we deduce that there exist leaves l_1 and l_2 of T (possibly $l_1 = l_2$) for which the double inequality

$$c_{l_1}(S,T) \leq \frac{|S| \cdot c(S,T)}{|T|} \leq c_{l_2}(S,T) \tag{4.2.1}$$

is satisfied. By relation (4.2.1), the number of copies of S in T not involving the leaf l_1 is

$$c(S,T) - c_{l_1}(S,T) \geq \Big(1 - \frac{|S|}{|T|}\Big) c(S,T)\,.$$

Let T^- be the topological tree that results when the leaf l_1 of T is removed and the unique vertex adjacent to l_1 is erased if it has outdegree 2 in T. We have

$$c(S,T^-) \geq \Big(1 - \frac{|S|}{|T|}\Big) c(S,T)\,,$$

and dividing by $\binom{|T|-1}{|S|}$, we obtain

$$\gamma(S,T^-) \geq \gamma(S,T)\,.$$

Since T is an arbitrary topological tree, this implies that

$$\max_{\substack{|T|=n-1 \\ T \text{ topological tree}}} \gamma(S,T) \geq \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)$$

for every $n \geq 1 + |S|$. In particular, the assertion on the limit follows as the sequence

$$\Big(\max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T) \Big)_{n \geq |S|}$$

is nonincreasing and bounded below. Moreover, we get

$$J(S) \leq \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T) \tag{4.2.2}$$

for all $n \geq |S|$. Now, denote by T^+ the tree obtained after replacing the leaf l_2 of T by an internal vertex with two leaves l_2 and l_2' attached to it. So, $c(S,T)$ represents the number of copies of S in T^+ not involving l_2'

(removing the leaf l_2' yields T) whereas the number of copies of S in T^+ involving l_2' is at least $c_{l_2}(S,T)$. Thus, it follows from relation (4.2.1) that

$$c(S,T^+) \geq \Big(1 + \frac{|S|}{|T|}\Big) c(S,T)\,.$$

Dividing by $\binom{|T|+1}{|S|}$ yields

$$\gamma(S,T^+) \geq \Big(1 - \frac{|S|(-1+|S|)}{|T|(|T|+1)}\Big)\gamma(S,T)\,,$$

and since T was assumed to be an arbitrary n-leaf topological tree, we have

$$\max_{\substack{|T|=n+1 \\ T \text{ topological tree}}} \gamma(S,T) \geq \Big(1 - \frac{|S|(-1+|S|)}{n(n+1)}\Big) \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)$$

for every $n \geq |S|$. After p iterations, we establish that

$$\max_{\substack{|T|=n+p \\ T \text{ topological tree}}} \gamma(S,T) \geq$$
$$\Big(\max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)\Big) \prod_{j=0}^{p-1} \Big(1 - \frac{|S|(-1+|S|)}{(n+p-j)(n+p-j-1)}\Big)$$

for all n, p with $p \geq 1$ and $n \geq |S|$. Letting $p \to \infty$ gives us the estimate

$$\begin{aligned} J(S) &\geq \Big(\max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)\Big)\Big(1 - \sum_{i=0}^{\infty} \frac{|S|(-1+|S|)}{(n+i+1)(n+i)}\Big) \\ &= \Big(\max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)\Big)\Big(1 - \frac{|S|(-1+|S|)}{n}\Big) \end{aligned} \tag{4.2.3}$$

for every $n \geq |S|$, where in the first step we used the standard inequality (which is proved by a simple induction) $\prod_{i=1}^{m}(1-y_i) \geq 1 - \sum_{i=1}^{m} y_i$ (valid when $0 \leq y_i \leq 1$ for every i – see [47, p. 60]) giving us

$$\prod_{j=0}^{p-1}\Big(1 - \frac{|S|(-1+|S|)}{(n+p-j)(n+p-j-1)}\Big) \geq 1 - \sum_{j=0}^{p-1} \frac{|S|(-1+|S|)}{(n+p-j)(n+p-j-1)}\,.$$

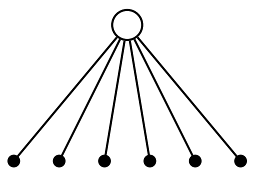
Putting relations (4.2.2) and (4.2.3) together, we obtain the desired double inequality and in particular, the asymptotic formula

$$\max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T) = J(S) + \mathcal{O}(n^{-1})$$

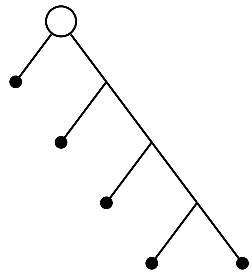
which is valid for all $n \geq |S|$. This completes the proof of the theorem. □

The problem of determining extremal graph structures that maximise or minimise a given graph parameter is a topical subject within graph theory. In what follows, we address the question of characterising all the topological trees that have the maximal inducibility 1. Unlike the degree-restricted inducibility $I_d(S)$, two families of topological trees are found to be maximal with respect to $J(S)$.

We recall that a binary caterpillar is a binary tree whose non-leaf vertices lie on a single path starting at the root, while a star is a topological tree in which all leaves are children of the root. The binary caterpillar with k leaves is denoted by F_k^2, and the star with k leaves is denoted by C_k. An illustration of both classes of trees is given in Figure 4.1.



The star C_6 with six leaves.



The binary caterpillar F_5^2 with five leaves.

Figure 4.1: A star and a binary caterpillar.

Theorem 4.2.0.2. Both stars and binary caterpillars have the maximal inducibility 1. Moreover, every topological tree S that is not a star or a binary caterpillar satisfies $J(S) < 1$.

We remark that Theorem 4.2.0.2 has a link to an observation made by Bubeck and Linial in [29]. In their context, the inducibility $\text{ind}(R)$ of a tree R (not necessarily a topological tree) with k vertices is defined to be the limit superior of the proportion of R as a subtree among all k-vertex induced subtrees where the limit is taken over all sequences of trees whose number of vertices grows to infinity. In the last section of their paper, Bubeck and Linial point out that stars and paths are the only trees R that have an inducibility $\text{ind}(R)$ of 1.

Proof of Theorem 4.2.0.2. First off, note that we have both $J(C_k) = 1$ and $J(F_k^2) = 1$ for every k in view of the identity

$$c(C_k, C_n) = c(F_k^2, F_n^2) = \binom{n}{k},$$

i.e., every subset of k leaves of the star C_n induces C_k, and every subset of k leaves of the binary caterpillar F_n^2 induces F_k^2.

The rest of the proof is a refinement of the proof of Theorem 3.6.0.1 in Chapter 3. To begin, let us prove that for any fixed positive integers $k > 2$ and $n > (k-1)^{k-2}$, every n-leaf topological tree contains either a copy of the k-leaf star C_k or a copy of the k-leaf binary caterpillar F_k^2.

Note that a topological tree has C_k as a leaf-induced subtree if and only if it contains a vertex that has at least k children. Likewise, a topological tree contains F_k^2 as a leaf-induced subtree if and only if it has height at least $k-1$. Fix $k > 2$, $n > (k-1)^{k-2}$, and consider a topological tree T with n leaves. If we suppose that T is a $(k-1)$-ary tree and T has height at most $k-2$, then T must have at most $(k-1)^{k-2}$ leaves because it is easy to see that for any fixed integers $h \geq 0$ and $m \geq 2$, an m-ary ($m \geq 2$) tree with height at most h can never have more than m^h leaves. This contradicts our assumption that $|T| = n > (k-1)^{k-2}$. Therefore, for $k > 2$ and $n > (k-1)^{k-2}$, every n-leaf topological tree must contain either a copy of the star C_k or a copy of the binary caterpillar F_k^2.

As a next step, if S is neither $F_{|S|}^2$ nor $C_{|S|}$, then by the discussion in the preceding paragraph, we obtain $c(S,T) < \binom{|T|}{|S|}$ for every topological tree with n leaves if $n > (|S|-1)^{|S|-2}$. Therefore, fixing $n > (|S|-1)^{|S|-2}$, we get

$$\max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T) < 1\,.$$

Furthermore, we know from the proof of Theorem 4.2.0.1 that

$$J(S) \leq \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)$$

for every $n \geq |S|$: this completes the proof of the theorem. $\square$

In particular, one obtains:

Corollary 4.2.0.3. We have

$$\liminf_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \gamma(S,T) = 0$$

for every topological tree S with at least three leaves.

Proof. This is a consequence of the identities $J(C_k) = 1$ and $J(F_k^2) = 1$ because

$$\liminf_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \gamma(S,T) \leq \liminf_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \left(1 - \gamma(F_{|S|}^2, T)\right) = 1 - J(F_{|S|}^2)$$

for every topological tree S different from a binary caterpillar, and likewise

$$\liminf_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \gamma(S,T) \leq \liminf_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \left(1 - \gamma(C_{|S|}, T)\right) = 1 - J(C_{|S|})$$

for every topological tree S different from a star. $\square$

As one might expect (in view of the aforementioned analogy to the work [29]), the quantity

$$\liminf_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \left(\gamma(C_k,T) + \gamma(F_k^2,T)\right)$$

is positive for every k. It seems arduous to determine its explicit value as a function of k. Nevertheless, we are able to say more about this quantity.

In fact, Bubeck and Linial [29] proved that in the context of induced subtrees of trees, the sum of the proportions of the k-vertex path and the k-vertex star is always greater than zero in the limit for every k. Specifically, they showed that the sum of the two proportions is bounded from below by an explicit constant that depends solely on k.

Proposition 4.2.0.4. For every fixed positive integer $k \geq 2$, we have

$$\begin{aligned} \liminf_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \left(\gamma(C_k,T) + \gamma(F_k^2,T)\right) &\geq \frac{1}{\binom{1+(k-1)^{k-2}}{k}} \\ &\geq \frac{k!}{\left(1+(k-1)^{k-2}\right)^k}. \end{aligned}$$

Proof. Let $k \geq 2$ be an arbitrary but fixed positive integer. In analogy to the proof of Theorem 4.2.0.1, we have

$$\sum_{l \in L(T)} \big(c_l(F_k^2, T) + c_l(C_k, T)\big) = k\big(c(F_k^2, T) + c(C_k, T)\big)$$

for every topological tree T. So there exists $l_1 \in L(T)$ such that the inequality

$$c_{l_1}(F_k^2, T) + c_{l_1}(C_k, T) \geq \frac{k}{|T|}\big(c(F_k^2, T) + c(C_k, T)\big)$$

holds. If we denote by T^- the topological tree obtained when removing the leaf l_1 of T as well as erasing the unique vertex adjacent to l_1 in T if it has outdegree 2 in T, then this implies that

$$\begin{aligned} c(F_k^2, T^-) + c(C_k, T^-) &= \big(c(F_k^2, T) + c(C_k, T)\big) - \big(c_{l_1}(F_k^2, T) + c_{l_1}(C_k, T)\big) \\ &\leq \Big(1 - \frac{k}{|T|}\Big)\big(c(F_k^2, T) + c(C_k, T)\big), \end{aligned}$$

and dividing by $\binom{|T|-1}{k}$ yields

$$\gamma(F_k^2, T^-) + \gamma(C_k, T^-) \leq \gamma(F_k^2, T) + \gamma(C_k, T)$$

for every topological tree T with $n \geq k$ leaves. Since T is arbitrary, it follows that

$$\min_{\substack{|T|=n-1 \\ T \text{ topological tree}}} \big(\gamma(F_k^2, T) + \gamma(C_k, T)\big) \leq \min_{\substack{|T|=n \\ T \text{ topological tree}}} \big(\gamma(F_k^2, T) + \gamma(C_k, T)\big),$$

and so the sequence

$$\Bigg(\min_{\substack{|T|=n \\ T \text{ topological tree}}} \big(\gamma(C_k, T) + \gamma(F_k^2, T)\big)\Bigg)_{n \geq k}$$

is nondecreasing. Using this result, we derive that

$$\min_{\substack{|T|=n \\ T \text{ topological tree}}} \big(\gamma(C_k, T) + \gamma(F_k^2, T)\big) \geq \min_{\substack{|T|=1+(k-1)^{k-2} \\ T \text{ topological tree}}} \big(\gamma(C_k, T) + \gamma(F_k^2, T)\big)$$

for every $n > (k-1)^{k-2}$. It follows that

$$\liminf_{\substack{|T| \to \infty \\ T \text{ topological tree}}} \big(\gamma(C_k, T) + \gamma(F_k^2, T)\big) \geq \min_{\substack{|T|=1+(k-1)^{k-2} \\ T \text{ topological tree}}} \big(\gamma(C_k, T) + \gamma(F_k^2, T)\big).$$

Furthermore, we know from the proof of Theorem 4.2.0.2 that

$$\gamma(C_k, T) + \gamma(F_k^2, T) \geq \frac{1}{\binom{|T|}{k}}$$

as soon as $|T| > (k-1)^{k-2}$. Hence, we obtain

$$\liminf_{\substack{|T| \to \infty \\ T \text{ topological tree}}} \left(\gamma(C_k, T) + \gamma(F_k^2, T)\right) \geq \frac{1}{\binom{1+(k-1)^{k-2}}{k}} .$$

This completes the proof of the proposition. $\square$

We would very much welcome seeing a solution to the following question:

QUESTION 4.2.0.5. What is the precise value of

$$\liminf_{\substack{|T| \to \infty \\ T \text{ topological tree}}} \left(\gamma(C_k, T) + \gamma(F_k^2, T)\right)$$

when $k > 3$?

Our next result gives a simple identity involving $J(S)$ and its d-ary counterpart $I_d(S)$. Specifically, although using equation (4.1.1) can be very difficult, we discover that to compute the inducibility $J(S)$, it is enough to properly understand the inducibility $I_d(S)$ as a single variable function of d.

Theorem 4.2.0.6. For every topological tree S,

$$J(S) = \lim_{d \to \infty} I_d(S) .$$

Proof. Let $\epsilon > 0$ be an arbitrary but fixed positive real number. We shall make use of the two estimates

$$J(S) \leq \max_{\substack{|T| = n \\ T \text{ topological tree}}} \gamma(S, T) \tag{4.2.4}$$

and

$$I_d(S) \geq \left(1 - \frac{|S|(-1 + |S|)}{n}\right) \max_{\substack{|T| = n \\ T \ d\text{-ary tree}}} \gamma(S, T)$$

both valid for every $n \geq |S|$. They follow from the proof of Theorem 4.2.0.1 and Theorem 3.3.0.1 in Chapter 3, respectively. For any positive integer $m_\epsilon > |S|(-1+|S|)/\epsilon$, we have

$$\frac{|S|(-1+|S|)}{m_\epsilon} \cdot \max_{\substack{|T|=m_\epsilon \\ T\ m_\epsilon-\text{ary tree}}} \gamma(S,T) < \epsilon\,,$$

which implies

$$\max_{\substack{|T|=m_\epsilon \\ T\ \text{topological tree}}} \gamma(S,T) - \Big(1 - \frac{|S|(-1+|S|)}{m_\epsilon}\Big) \max_{\substack{|T|=m_\epsilon \\ T\ m_\epsilon-\text{ary tree}}} \gamma(S,T) < \epsilon\,. \tag{4.2.5}$$

Employing the relation

$$I_{m_\epsilon}(S) \geq \Big(1 - \frac{|S|(-1+|S|)}{m_\epsilon}\Big) \max_{\substack{|T|=m_\epsilon \\ T\ m_\epsilon-\text{ary tree}}} \gamma(S,T)\,,$$

and invoking the fact that $\big(I_m(S)\big)_{m\geq 2}$ is a nondecreasing sequence of real numbers, inequality (4.2.5) implies

$$\Big(\max_{\substack{|T|=m_\epsilon \\ T\ \text{topological tree}}} \gamma(S,T)\Big) - I_m(S) < \epsilon\,,$$

for every $m \geq m_\epsilon$. Combining this with (4.2.4), we establish that

$$J(S) - I_m(S) < \epsilon$$

for every $m \geq m_\epsilon$. On the other hand, we have

$$J(S) \geq \sup_{d\geq 2} I_d(S) = \lim_{d\to\infty} I_d(S)$$

by definition of $J(S)$. Hence, we conclude that for every $\epsilon > 0$, there exists $\Delta \geq 2$ such that the inequality

$$|J(S) - I_d(S)| < \epsilon$$

holds for every $d > \Delta$. This completes the proof of the theorem. □

Remark 4.2.0.7. For a topological tree T, let us denote by $\Delta(T)$ the maximum among the outdegrees of the vertices of T. Let $T_1^\bullet, T_2^\bullet, \ldots$ be a sequence of topological trees such that $|T_n^\bullet| \to \infty$ as $n \to \infty$ and

$$J(S) = \lim_{n\to\infty} \gamma(S, T_n^\bullet)\,.$$

If $\max\limits_{i\geq 1} \Delta(T_i^\bullet)$ exists, then the sequence $\big(I_d(S)\big)_{d\geq 2}$ becomes constant from the point $\Delta = \max\limits_{i\geq 1} \Delta(T_i^\bullet)$ onwards, that is,

$$I_\Delta(S) = I_{1+\Delta}(S) = I_{2+\Delta}(S) = \cdots\,.$$

As already mentioned earlier, Theorem 4.2.0.6 can be very useful. The following corollary reveals one of its features:

Corollary 4.2.0.8. We have $J(B) = I_2(B)$ for every binary tree B.

Proof. It is shown (Theorem 3.4.0.1) in Chapter 3 that $I_d(B) = I_2(B)$ for every d and binary tree B. Therefore, by Theorem 4.2.0.6, we obtain $J(B) = I_2(B)$. $\square$

For the next corollary, we first need to recall the definition of a complete d-ary tree. The complete d-ary tree CD_h^d of height h is the d-ary tree defined recursively in the following way: CD_0^d is the single leaf; for $h > 0$, CD_h^d has d branches isomorphic to the tree CD_{h-1}^d. In other words, a complete d-ary tree is a d-ary tree in which every internal vertex has outdegree d and all the leaves (d^h in total) are at the same distance h from the root.

Corollary 4.2.0.9. There are infinitely many topological trees S with $J(S) \leq \epsilon$ for every $\epsilon > 0$.

Proof. It is sufficient to consider binary trees. By a result that appears in [37] (see Corollary 5.3.0.4 in Chapter 5), we know that the inducibility $I_2(CD_h^2)$ of the complete binary tree CD_h^2 of height h is at most $I_2(CD_2^2)^{2^{h-2}}$ for all $h \geq 2$. Therefore, since it is also proved in the same Chapter 5 (see Theorem 5.4.0.2 and also paper [1]) that $I_2(CD_2^2) = 3/7$, we deduce that

$$I_2(CD_h^2) \leq I_2(CD_2^2)^{2^{h-2}} = \left(\frac{3}{7}\right)^{2^{h-2}} \to 0 \;\text{ as }\; h \to \infty\,.$$

Thus, the assertion of the corollary follows from Corollary 4.2.0.8. $\square$

We close this section by collecting two general facts about the inducibility $J(S)$. We note that some results in Chapter 3 with respect to $I_d(S)$ hold analogously for $J(S)$ using our Theorem 4.2.0.6. Note that $J(S) > 0$ for every topological tree S because by definition, we have $J(S) \geq I_{\Delta(S)}(S)$ where $\Delta(S)$ is the maximum outdegree among the vertices of S while it is proved in Chapter 3 that $I_{\Delta(S)}(S) > 0$.

The following proposition provides a better lower bound that only depends on the number of leaves of S.

Proposition 4.2.0.10. We have

$$J(S) \geq \frac{(-1+|S|)!}{-1+|S|^{-1+|S|}}$$

for every topological tree S with at least two leaves.

Proof. Consider a topological tree S. By a result in Chapter 3, we have

$$I_d(S) \geq \frac{(-1+|S|)!}{-1+|S|^{-1+|S|}}$$

for every $d \geq \Delta(S)$. Passing to the limit as $d \to \infty$, Theorem 4.2.0.6 gives us the desired statement. □

For two topological trees S_1, S_2, denote by $\mathcal{F}(S_1; S_2)$ the unique topological tree which is constructed by appending the root of S_2 to every leaf of S_1. See Figure 4.2 for a diagram.

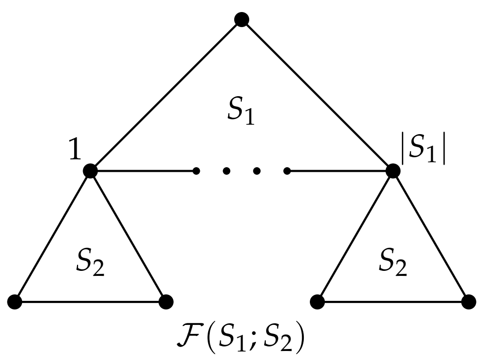


Figure 4.2: The shape of the tree $\mathcal{F}(S_1; S_2)$ defined for Theorem 4.2.0.11.

Theorem 4.2.0.11. The tree $\mathcal{F}(S_1; S_2)$ satisfies

$$J\big(\mathcal{F}(S_1; S_2)\big) \geq \frac{(|S_1| \cdot |S_2|)!}{(|S_2|!)^{|S_1|} \cdot |S_1|^{|S_1| \cdot |S_2|}} \cdot J(S_2)^{|S_1|}$$

for every pair of topological trees S_1 and S_2.

Proof. Again, by Theorem 3.5.0.2 in Chapter 3, we have

$$I_d\big(\mathcal{F}(S_1; S_2)\big) \geq \frac{(|S_1| \cdot |S_2|)!}{(|S_2|!)^{|S_1|} \cdot |S_1|^{|S_1| \cdot |S_2|}} \cdot I_d(S_2)^{|S_1|}$$

for every $d \geq \Delta\big(\mathcal{F}(S_1; S_2)\big)$. Taking the limit of both sides as $d \to \infty$, and invoking Theorem 4.2.0.6, we obtain the desired inequality of the theorem. □

Let us call a topological tree *full* if for every internal vertex v, all the branches of the subtree rooted at v are isomorphic. An example of a full topological tree is shown in Figure 4.3.

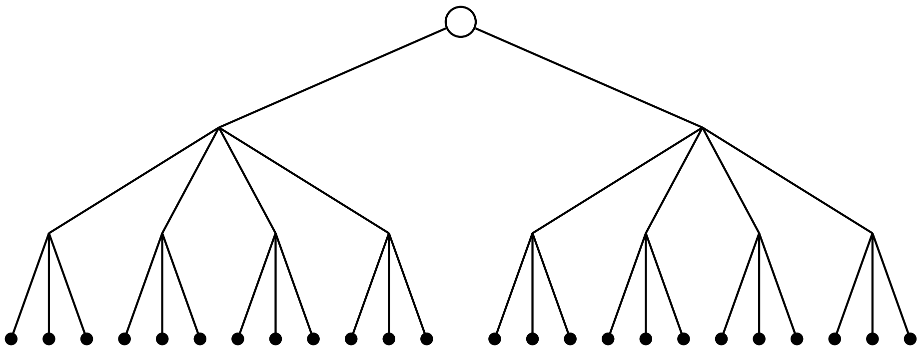


Figure 4.3: A full topological tree.

Equivalently, a topological tree S is full if and only if all of its leaf-induced subtrees with $|S| - 1$ leaves are isomorphic. With this observation, we produce an analogue of Proposition 4.2.0.10 for such a leaf-induced subtree of a full topological tree. In this special case, we obtain a better result:

Proposition 4.2.0.12. Let T be an arbitrary full d-ary tree (with at least three leaves) and $\hat{D}$ its leaf-induced subtree with $|T| - 1$ leaves. Then we have

$$I_d(\hat{D}) \geq \frac{|\hat{D}|!}{-1 + (1 + |\hat{D}|)^{-1+|\hat{D}|}}.$$

In particular, one obtains

$$J(\hat{D}) \geq \frac{|\hat{D}|!}{-1+(1+|\hat{D}|)^{-1+|\hat{D}|}}.$$

Proof. The proof is a refinement of the proof of Proposition 3.5.0.1 in Chapter 3. Fix $d \geq 2$ and a full d-ary tree T with at least three leaves. Set $T^{[1]} = T$ and then define recursively the family $T^{[n+1]} = \mathcal{F}(T; T^{[n]})$ of strictly d-ary trees. So $|T^{[n]}| = |T|^n$ for every $n \geq 1$.

Let $\hat{D}$ be the unique leaf-induced subtree of T with $|T|-1$ leaves. There are $|T| \cdot c(\hat{D}, T^{[n]})$ copies of $\hat{D}$ in the $|T|$ copies of $T^{[n]}$ that are attached to the leaves of T to obtain $T^{[n+1]}$. Also, if one chooses $|T|-1$ of these $|T|$ copies and takes one arbitrary leaf from each of them, one obtains a total of another $|T| \cdot |T^{[n]}|^{|T|-1}$ copies of $\hat{D}$. This yields a lower bound on $c(\hat{D}, T^{[1+n]})$, namely

$$c(\hat{D}, T^{[1+n]}) \geq |T| \cdot c(\hat{D}, T^{[n]}) + |T| \cdot |T^{[n]}|^{|T|-1},$$

and by iteration (starting with $c(\hat{D}, T^{[0]}) = 0$), we get

$$\begin{aligned} c(\hat{D}, T^{[1+n]}) &\geq \sum_{j=0}^{n} |T|^{1+j} \cdot |T^{[-j+n]}|^{|T|-1} \\ &= |T|^{(n+1)\cdot(|T|-1)} \cdot |T|^{2-|T|} \cdot \frac{1-|T|^{(n+1)\cdot(2-|T|)}}{1-|T|^{2-|T|}}. \end{aligned}$$

Since $|T| \geq 3$, it follows that

$$\begin{aligned} I_d(\hat{D}) \geq \limsup_{n\to\infty} \gamma\big(\hat{D}, T^{[n]}\big) &\geq (|T|-1)! \cdot \frac{|T|^{2-|T|}}{1-|T|^{2-|T|}} \\ &= \frac{|\hat{D}|!}{-1+(1+|\hat{D}|)^{-1+|\hat{D}|}}, \end{aligned}$$

proving the assertion of the first part of the proposition. Moreover, we have

$$I_m(\hat{D}) \geq \frac{|\hat{D}|!}{-1+(1+|\hat{D}|)^{-1+|\hat{D}|}},$$

for all $m \geq d$. So passing to the limit as $m \to \infty$ and invoking Theorem 4.2.0.6, we obtain the second assertion of the proposition. $\square$

It turns out that if one knows the inducibility of a topological tree S, then one can obtain a lower bound on the inducibility of any of the leaf-induced subtrees of S:

Theorem 4.2.0.13. For any three topological trees R, S and T such that $|T| \geq |S| \geq |R|$, we have

$$c(R,T) \geq \frac{c(R,S)}{\binom{|T|-|R|}{|S|-|R|}} \cdot c(S,T)\,.$$

In particular, we obtain

$$J(R) \geq J(S)\gamma(R,S)\,.$$

Proof. Given a topological tree S, we can count the number of appearances of any smaller topological tree R in a larger topological tree T by first counting the number of copies of S in T and then the number of copies of R in S. Clearly, $c(R,T)$ is overcounted in this way based on the observation that the intersection of the subsets of leaves of T that induce a copy of S may not be empty. So we would like to take this observation into account.

Assume $|T| \geq |S| \geq |R|$. Given a subset L of leaves of T that induces a copy of R, we can then choose any subset L' of $|S| - |R|$ leaves from the set of leaves of T without L so that $L \cup L'$ induces a subtree of T with $|S|$ leaves. In particular, all copies of S are obtained $c(R,S)$ times in this way. We deduce that the quantity $c(S,T) \cdot c(R,S)$ is at most $\binom{|T|-|R|}{|S|-|R|} \cdot c(R,T)$. Consequently, one obtains a simple lower bound on $c(R,T)$, namely

$$c(R,T) \geq \frac{c(S,T) \cdot c(R,S)}{\binom{|T|-|R|}{|S|-|R|}}\,.$$

As a next step, we take the density:

$$\begin{aligned}\gamma(R,T) &\geq \gamma(R,S) \cdot \frac{\binom{|S|}{|R|} \cdot \binom{|T|}{|S|}}{\binom{|T|}{|R|} \cdot \binom{|T|-|R|}{|S|-|R|}} \cdot \gamma(S,T) \\ &= \gamma(R,S) \cdot \gamma(S,T)\,.\end{aligned}$$

Finally, in view of Theorem 4.2.0.1, we can take the limit to obtain the desired result. $\square$

Remark 4.2.0.14. Let $d \geq 2$ be a fixed positive integer. By the same argument as in the proof of Theorem 4.2.0.13, it is also seen that

$$I_d(D_1) \geq I_d(D_2) \cdot \gamma(D_1, D_2)$$

for any two d-ary trees D_1 and D_2 satisfying $|D_1| \leq |D_2|$.

4.3 Concluding comments

It would be interesting to answer the following question:

QUESTION 4.3.0.1. Can we explicitly determine the inducibility $J(S)$ of any topological tree S other than a star or a binary tree? Note that in the degree-restricted context, the answer to this question is affirmative [37] (see Chapter 5).

Finally, we conjecture that equality never holds in Proposition 4.2.0.10. It is also natural to formulate the following problem:

QUESTION 4.3.0.2. Is there a sequence $(T_n)_{n \geq 1}$ of topological trees such that $|T_n| = n$ and $\lim_{n \to \infty} \gamma(S, T_n)$ exists and is positive for all topological trees S? In the degree-restricted context, the answer is positive [37]—see Chapter 5.

Chapter 5

Further results on the inducibility of d-ary trees

We give a general upper bound on the inducibility of D as a function of the inducibilities of its branches. Moreover, we demonstrate that the bound is sharp for infinitely many d-ary trees. A d-ary tree is called *balanced* if the number of leaves of its branches pairwise differ at most by one. We obtain an improved upper bound on the inducibility of an arbitrary balanced d-ary tree. We give several examples proving that the bound is sharp for every given number of leaves. In particular, the precise inducibilities of certain balanced d-ary trees are derived. Furthermore, we present a lower bound that asymptotically matches the (improved) upper bound under specific restrictions. We also demonstrate that the sequence of complete d-ary trees contains a positive density of any fixed d-ary tree in the limit.

The material in this chapter will appear as the following paper [37]: *Further Results on the Inducibility of d-ary Trees. A. A. V. Dossou-Olory, and S. Wagner. To be submitted.*

5.1 Introduction

Broadly speaking, the inducibility of a tree provides a measure of the largest density at which the given tree can be found in a tree whose size gets bigger. Recall that, the inducibility $I_d(D)$ of a d-ary tree D is the limit superior, taken over all d-ary trees, of the density of subsets of $|D|$ leaves of T that induce a copy of D:

$$I_d(D) := \limsup_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \frac{c(D,T)}{\binom{|T|}{|D|}}.$$

From now on, unless otherwise specified, d is always an arbitrary but fixed positive integer greater than 1.

Set $\gamma(D,T) = c(D,T)/\binom{|T|}{|D|}$, which is the probability that $|D|$ distinct random leaves of T induce a copy of D, or the density of D in T (for short). In Chapter 3, we proved the following two asymptotic formulas:

$$\begin{aligned}\max_{\substack{|T|=n \\ T\ \text{strictly}\ d\text{-ary tree}}} \gamma(D,T) &= I_d(D) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right), \\ \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T) &= I_d(D) + \mathcal{O}\left(\frac{1}{n}\right)\end{aligned} \tag{5.1.1}$$

and hence,

$$I_d(D) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T\ \text{strictly}\ d\text{-ary tree}}} \gamma(D,T).$$

Moreover, it was shown, among other things, that $I_d(D) > 0$ for every d-ary tree D, and we also gave lower bounds on the inducibility $I_d(D)$ under various assumptions. The purpose of this chapter is multifold:

- we develop a new general lower bound on the inducibility of an arbitrary d-ary tree—Theorem 5.3.0.1;
- we establish a general upper bound on the inducibility of a d-ary tree as a function of the inducibilities of its branches—Theorem 5.3.0.2 (and Corollary 5.3.0.5 for instance);
- we present an improved upper bound on the inducibilities of the d-ary trees that are 'balanced' as a function of the inducibilities of their branches—Theorem 5.4.0.2;
- we give several examples showing that the improved upper bound is also sharp, and furthermore, provide a characterisation of the d-ary trees that attain the bound (Theorem 5.4.0.7);

- we give an asymptotic formula for the density of certain balanced d-ary trees in strictly d-ary trees (Theorem 5.4.0.7)—in this special case, this improves on the error term in (5.1.1); we recall that it is an open question whether the bound $\mathcal{O}(|T|^{-1/2})$ can be improved further for all d-ary trees D and every $d > 2$.

5.2 Setting up a general recursion

Rooted trees are predestined for recursive approaches. For a d-ary tree D with branches $D_1, D_2, \ldots, D_r$, we define the equivalence relation $\sim_D$ on the set of all permutations of the indices $1, 2, \ldots, r$ as follows: for two permutations π and π' of $\{1, 2, \ldots, r\}$,

$$\big(\pi(1), \pi(2), \ldots, \pi(r)\big) \sim_D \big(\pi'(1), \pi'(2), \ldots, \pi'(r)\big)$$

if for every $j \in \{1, 2, \ldots, r\}$, the tree $D_{\pi(j)}$ is isomorphic (in the sense of rooted trees) to the tree $D_{\pi'(j)}$. Further, we denote by $M(D)$ a complete set of representatives of all equivalence classes of the equivalence relation $\sim_D$. Thus, if $m_1, m_2, \ldots, m_c$ denote the multiplicities of the branches of D with respect to the equivalence relation $\sim_D$, then the size of $M(D)$ is precisely

$$|M(D)| = \binom{r}{m_1, m_2, \ldots, m_c}.$$

We can set up a recursion for the number $c(D, T)$ of copies of D in a d-ary tree T. To this end, we make use of the set $M(D)$ which accounts for the possibility that some of the branches of D are isomorphic. We get the following identity:

$$c(D, T) = \sum_{i=1}^{d} c(D, T_i) + \sum_{\{i_1, i_2, \ldots, i_r\} \subseteq \{1, 2, \ldots, d\}} \ \sum_{\pi \in M(D)} \ \prod_{j=1}^{r} c\big(D_{\pi(j)}, T_{i_j}\big), \quad (5.2.1)$$

which is valid for every d-ary tree T with branches $T_1, T_2, \ldots, T_d$ (some branches are allowed to be empty). The proof of this formula is straightforward. In words, (5.2.1) is established as follows:

- The term $\sum_{i=1}^{d} c(D, T_i)$ is the number of subsets of leaves that belong to a single branch of T and induce a copy of D.

- The expression $\prod_{j=1}^{r} c\big(D_{\pi(j)}, T_{i_j}\big)$ stands for the number of copies of D in which its branches $D_{\pi(1)}, D_{\pi(2)}, \ldots, D_{\pi(r)}$ are induced by subsets of leaves of $T_{i_1}, T_{i_2}, \ldots, T_{i_r}$, respectively. We sum this expression over all subsets of r elements of the set of branches of T and all permutations π in $M(D)$, so as to take into consideration the possibility that some branches of D might be isomorphic.

Equation (5.2.1) will be used repeatedly in various places of this chapter.

Finally, we say that a sequence $T^1, T^2, \ldots$ of d-ary trees such that the number of leaves of T^n tends to infinity as $n \to \infty$, is asymptotically maximal for a d-ary tree D if

$$\lim_{n\to\infty} \gamma(D, T^n) = I_d(D)\,.$$

In other words, the sequence $T^1, T^2, \ldots$ of d-ary trees yields the inducibility of D in the limit.

5.3 Bounding the inducibility

For our first result, which offers a lower bound on the inducibility of an arbitrary d-ary tree, we need to define a specific class of d-ary trees. By a star, we mean a topological tree in which all its edges are incident with a single vertex (the root of the tree). The symbol C_k denotes the star with k leaves. The complete d-ary tree of height h – which we denote by CD_h^d – is defined recursively as follows: CD_0^d has only one vertex and for $h > 0$, the tree CD_h^d is obtained by joining d copies of CD_{h-1}^d (their respective roots) to a new common vertex (the root of the tree CD_h^d). Note that the height of a rooted tree is the distance from the root to a leaf farthest from the root and thus CD_h^d has precisely d^h leaves. For example, CD_1^d is the d-leaf star.

Our first two theorems demonstrate the special role that complete d-ary trees play in the study of the inducibility of certain d-ary trees. Moreover, the first theorem also infers that every d-ary tree appears in a positive density as a leaf-induced subtree in complete d-ary trees of sufficiently large height.

Theorem 5.3.0.1. The limit

$$\lim_{h\to\infty} \gamma(D, CD_h^d)$$

exists for every d-ary tree D and is given by

$$|M(D)| \binom{d}{r} \frac{\binom{|D|}{|D_1|,|D_2|,\ldots,|D_r|}}{d^{|D|} - d} \prod_{i=1}^{r} \lim_{h\to\infty} \gamma(D_i, CD_h^d),$$

where $D_1, D_2, \ldots, D_r$ denote the branches of D.

Proof. Consider a d-ary tree D with branches $D_1, D_2, \ldots, D_r$. We employ the normalised recurrence relation

$$\gamma(D, CD_h^d) = d \cdot \frac{\binom{d^{h-1}}{|D|}}{\binom{d^h}{|D|}} \cdot \gamma(D, CD_{h-1}^d)$$
$$+ \binom{d}{r} \sum_{\pi \in M(D)} \frac{\prod_{i=1}^{r} \binom{d^{h-1}}{|D_{\pi(i)}|}}{\binom{d^h}{|D|}} \prod_{i=1}^{r} \gamma(D_{\pi(i)}, CD_{h-1}^d)$$

obtained through the specialisation $T = CD_h^d$ in the general recursion (5.2.1) when passing to the density $\gamma(D, T)$ (recall that for $h > 0$, all the branches of CD_h^d are isomorphic to CD_{h-1}^d).

Letting $h \to \infty$, and applying $\liminf$ to both sides of this normalised equation, we get

$$\liminf_{h\to\infty} \gamma(D, CD_h^d) \geq d^{1-|D|} \cdot \liminf_{h\to\infty} \gamma(D, CD_{h-1}^d)$$
$$+ \binom{d}{r} \sum_{\pi \in M(D)} \Bigg[\frac{|D|!}{|D_{\pi(1)}|! \cdot |D_{\pi(2)}|! \cdot \ldots \cdot |D_{\pi(r)}|!} \cdot$$
$$\left(\prod_{i=1}^{r} d^{-|D_{\pi(i)}|} \right) \prod_{i=1}^{r} \liminf_{h\to\infty} \gamma(D_{\pi(i)}, CD_{h-1}^d) \Bigg],$$

which implies (after rearranging terms accordingly) that

$$\liminf_{h\to\infty} \gamma(D, CD_h^d) \geq$$
$$|M(D)| \binom{d}{r} \frac{\binom{|D|}{|D_1|,|D_2|,\ldots,|D_r|}}{d^{|D|} - d} \prod_{i=1}^{r} \liminf_{h\to\infty} \gamma(D_i, CD_h^d)$$

as $|D_1| + |D_2| + \cdots + |D_r| = |D|$. In the same manner, now using lim sup as $h \to \infty$, we also obtain

$$\limsup_{h\to\infty} \gamma(D, CD_h^d) \leq$$
$$|M(D)| \binom{d}{r} \frac{\binom{|D|}{|D_1|,|D_2|,\ldots,|D_r|}}{d^{|D|} - d} \prod_{i=1}^{r} \limsup_{h\to\infty} \gamma(D_i, CD_h^d)\,.$$

Hence, we can conclude that the desired statement of the theorem follows by induction on the height of D, starting with height 0 (in which case the statement is trivial). □

The lower bound on $I_d(D)$ derived in Theorem 5.3.0.1 is attained for every complete d-ary tree, for instance; see Theorem 5.4.0.2.

Our next result is a general upper bound on the inducibility of every d-ary tree. To be precise, the result is an explicit inequality between the inducibilities of a d-ary tree and its branches.

Theorem 5.3.0.2. Let D be a d-ary tree with branches $D_1, D_2, \ldots, D_r$. Then the inequality

$$I_d(D) \leq \binom{|D|}{|D_1|, |D_2|, \ldots, |D_r|} \Big(\prod_{i=1}^{r} I_d(D_i)\Big) \cdot$$
$$\sup_{\substack{0 \leq x_1, x_2, \ldots, x_d < 1 \\ x_1 + x_2 + \cdots + x_d = 1}} \left\{ \frac{1}{1 - \sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1, i_2, \ldots, i_r\} \subseteq \{1,2,\ldots,d\}} \sum_{\pi \in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|} \right\}$$

holds.

The bound in Theorem 5.3.0.2 can also be attained; see Corollary 5.3.0.5 for instance.

Another ingredient is needed in order to prove Theorem 5.3.0.2:

Lemma 5.3.0.3. Let D be a d-ary tree whose branches are $D_1, D_2, \ldots, D_r$. Assume that branches with the same number of leaves are isomorphic.

Then

$$\sup_{\substack{0\le x_1,x_2,\dots,x_d<1\\ x_1+x_2+\cdots+x_d=1}} \left\{ \frac{1}{1-\sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1,i_2,\dots,i_r\}\subseteq\{1,2,\dots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|} \right\} \le \frac{1}{\binom{|D|}{|D_1|,|D_2|,\dots,|D_r|}}\,.$$

Furthermore, if $r=2,\ |D_1|=1$ and $|D_2|>1$ then

$$\sup_{\substack{0\le x_1,x_2,\dots,x_d<1\\ x_1+x_2+\cdots+x_d=1}} \left\{ \frac{1}{1-\sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1,i_2\}\subseteq\{1,2,\dots,d\}} \left(x_{i_1}\cdot x_{i_2}^{|D|-1} + x_{i_1}^{|D|-1}\cdot x_{i_2} \right) \right\} = |D|^{-1}\,.$$

We defer the proof of the lemma to the end of the section and now prove Theorem 5.3.0.2:

Proof of Theorem 5.3.0.2. Let D be a d-ary tree whose branches are denoted by $D_1, D_2, \dots, D_r$. It is easy to see that for $|D|=2$, the inequality in the theorem holds with equality. In fact, for $|D|=2$,

$$\frac{1}{1-\sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1,i_2,\dots,i_r\}\subseteq\{1,2,\dots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|} = \frac{1}{1-\sum_{i=1}^{d} x_i^{2}} \sum_{\{i_1,i_2\}\subseteq\{1,2,\dots,d\}} x_{i_1}\cdot x_{i_2} = \frac{1}{2}$$

by virtue of the multinomial theorem, while we have $I_d(D_1)=I_d(D_2)=I_d(D)=1$. So we can assume that D has more than two leaves. We know from the proof of Theorem 3.3.0.1 in Chapter 3 that

$$0 \le \max_{\substack{|T|=n\\ T\ d\text{-ary tree}}} \gamma(D,T) - I_d(D) \le \frac{|D|(|D|-1)}{n} \tag{5.3.1}$$

for all $n \ge |D|$. Consider a sequence $T_1, T_2, \dots$ of d-ary trees such that $|T_n|\to\infty$ as $n\to\infty$ and

$$\max_{\substack{|T|=|T_n|\\ T\ d\text{-ary tree}}} c(D,T) = c(D,T_n)\,.$$

In particular, the sequence $T_1, T_2, \ldots$ is asymptotically maximal for D. Denote the branches of T_n by $T_{n,1}, T_{n,2}, \ldots, T_{n,d}$ (some branches are allowed to be empty). One can assume that $T_{n,1}$ is a branch with the largest number of leaves for every n. Set $\alpha_{n,i} := |T_{n,i}|/|T_n|$ for every $i \in \{1, 2, \ldots, d\}$ and n. Set $\beta_n = 1 - \alpha_{n,1}$. Since $0 < \beta_n \leq 1$, we have

$$1 - \sum_{i=1}^{d} \alpha_{n,i}^{|D|} \geq 1 - \alpha_{n,1}^{|D|} - \Big(\sum_{i=2}^{d} \alpha_{n,i}\Big)^{|D|}$$
$$= 1 - \alpha_{n,1}^{|D|} - (1 - \alpha_{n,1})^{|D|} = 1 - (1 - \beta_n)^{|D|} - \beta_n^{|D|}\,.$$

We distinguish two cases based on whether β_n is 'small' or 'large' in the limit.

Case 1: Suppose that β_n is bounded below by a positive constant δ as $n \to \infty$. In this case we have

$$1 - \sum_{i=1}^{d} \alpha_{n,i}^{|D|} \geq 1 - (1 - \beta_n)^{|D|} - \beta_n^{|D|} \geq 1 - (1 - \delta)^{|D|} - \delta^{|D|}$$

for all n. According to (5.2.1), a recursion for the number of copies of D in T_n is given by

$$c(D, T_n) = \sum_{i=1}^{d} c(D, T_{n,i}) \ + \sum_{\{i_1, i_2, \ldots, i_r\} \subseteq \{1,2,\ldots,d\}} \ \sum_{\pi \in M(D)} \prod_{j=1}^{r} c(D_{\pi(j)}, T_{n,i_j})\,.$$

Using (5.3.1), we obtain

$$I_d(D)\binom{|T_n|}{|D|} \leq c(D, T_n) \leq \sum_{i=1}^{d} \Big(I_d(D) + \frac{|D|(|D|-1)}{|T_{n,i}|}\Big)\binom{|T_{n,i}|}{|D|}$$
$$+ \sum_{\{i_1, i_2, \ldots, i_r\} \subseteq \{1,2,\ldots,d\}} \ \sum_{\pi \in M(D)} \prod_{j=1}^{r} \Big(I_d(D_{\pi(j)}) + \frac{|D_{\pi(j)}|(|D_{\pi(j)}|-1)}{|T_{n,i_j}|}\Big)\binom{|T_{n,i_j}|}{|D_{\pi(j)}|}\,,$$

which implies that

$$I_d(D)\frac{|T_n|^{|D|}}{|D|!} + \mathcal{O}(|T_n|^{|D|-1}) \leq \sum_{i=1}^{d} \Big(I_d(D)\frac{|T_{n,i}^{|D|}}{|D|!} + \frac{|T_{n,i}|^{|D|-1}}{(|D|-2)!}\Big)$$
$$+ \sum_{\{i_1, i_2, \ldots, i_r\} \subseteq \{1,2,\ldots,d\}} \ \sum_{\pi \in M(D)} \prod_{j=1}^{r} \Big(I_d(D_{\pi(j)})\frac{|T_{n,i_j}|^{|D_{\pi(j)}|}}{|D_{\pi(j)}|!} + N(T_{n,i_j}, D_{\pi(j)})\Big)\,,$$

where $N(T_{n,i_j}, D_{\pi(j)})$ is equal to $|T_{n,i_j}|^{|D_{\pi(j)}|-1}/(|D_{\pi(j)}|-2)!$ if $|D_{\pi(j)}| > 2$, and 0 otherwise. Consequently,

$$\Big(|T_n|^{|D|} - \sum_{i=1}^{d} |T_{n,i}|^{|D|}\Big) I_d(D) \leq$$

$$|D|! \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} I_d(D_{\pi(j)}) \frac{|T_{n,i_j}|^{|D_{\pi(j)}|}}{|D_{\pi(j)}|!} + \mathcal{O}(|T_n|^{|D|-1})$$

as $|T_{n,i_j}| < |T_n|$ for all $j \in \{1,2,\ldots,r\}$ and n. Dividing through by $|T_n|^{|D|}$, we get

$$\Big(1 - \sum_{i=1}^{d} \alpha_{n,i}^{|D|}\Big) I_d(D) \leq \frac{|D|!}{|D_1|!\cdot|D_2|!\cdot\ldots\cdot|D_r|!}$$

$$\cdot \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} I_d(D_{\pi(j)}) \alpha_{n,i_j}^{|D_{\pi(j)}|} + \mathcal{O}(|T_n|^{-1}) \,.$$

Now using the fact that $1 - \sum_{i=1}^{d} \alpha_{n,i}^{|D|}$ is bounded below by a positive constant as $n \to \infty$, we deduce that

$$\begin{aligned} I_d(D) \leq & \frac{|D|!}{|D_1|!\cdot|D_2|!\cdot\ldots\cdot|D_r|!} \Big(\prod_{i=1}^{r} I_d(D_i)\Big) \\ & \cdot \frac{\sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} \alpha_{n,i_j}^{|D_{\pi(j)}|}}{1 - \sum_{i=1}^{d} \alpha_{n,i}^{|D|}} + \mathcal{O}(|T_n|^{-1}) \\ \leq & \Big(\prod_{i=1}^{r} I_d(D_i)\Big) \binom{|D|}{|D_1|,|D_2|,\ldots,|D_r|} \\ & \cdot \sup_{\substack{0\leq x_1,x_2,\ldots,x_d<1 \\ x_1+x_2+\cdots+x_d=1}} \frac{\sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|}}{1 - \sum_{i=1}^{d} x_i^{|D|}} \\ & + \mathcal{O}(|T_n|^{-1}) \,. \end{aligned}$$

Finally, we take the limit as $n \to \infty$: this gives us the desired result.

Case 2: One can then assume (without loss of generality) that the limit of β_n is actually 0 as $n \to \infty$ (by considering a subsequence). Denote by $T_n \backslash T_{n,1}$ the subtree induced by the leaves of T_n that are not leaves of $T_{n,1}$.

Claim 1: We claim that the number of copies of D in T_n that involve more than one leaf of $T_n \backslash T_{n,1}$ is at most of order $\mathcal{O}(\beta_n^2 \cdot |T_n|^{|D|})$.

For the proof of the claim, note that by definition, the number of copies of D in T_n that involve more than one leaf of $T_n \backslash T_{n,1}$ is at most

$$\sum_{l=2}^{|D|} \binom{|T_n| - |T_{n,1}|}{l} \binom{|T_{n,1}|}{|D| - l}.$$

On the other hand, we have

$$\begin{aligned}
\sum_{l=2}^{|D|} \binom{|T_n| - |T_{n,1}|}{l} \binom{|T_{n,1}|}{|D| - l} &\leq \sum_{l=2}^{|D|} \frac{(|T_n| - |T_{n,1}|)^l}{l!} \cdot \frac{|T_{n,1}|^{|D|-l}}{(|D| - l)!} \\
&= |T_n|^{|D|}(1 - \alpha_{n,1})^2 \sum_{l=2}^{|D|} \frac{(1 - \alpha_{n,1})^{l-2} \alpha_{n,1}^{|D|-l}}{l!(|D| - l)!} \\
&\leq |T_n|^{|D|} \cdot \beta_n^2 \sum_{l=2}^{|D|} \frac{1}{l!(|D| - l)!}.
\end{aligned}$$

This completes the proof of the claim. It follows that the proportion of copies of D in T_n that involve more than one leaf of $T_n \backslash T_{n,1}$ is of order $\mathcal{O}(\beta_n^2)$ among all subsets of $|D|$ leaves of T_n.

Claim 2: Based on another result from Chapter 3, we further claim that D must have only two branches, one of which is a single leaf.

Indeed, suppose that D does not have this shape. Then the subsets of leaves of T_n that induce a copy of D come in two varieties: either the $|D|$ leaves are all leaves of $T_{n,1}$, or more than one of the $|D|$ leaves is a leaf of $T_n \backslash T_{n,1}$. So this gives us

$$c(D, T_n) = c(D, T_{n,1}) + \mathcal{O}(\beta_n^2 \cdot |T_n|^{|D|}). \tag{5.3.2}$$

by Claim 1. It was established in the proof of (5.3.1) (Theorem 3.3.0.1) that

$$0 \leq \max_{\substack{|T'|=k \\ T' \text{ } d\text{-ary tree}}} \gamma(D, T') - \max_{\substack{|T''|=k+1 \\ T'' \text{ } d\text{-ary tree}}} \gamma(D, T'') \leq \frac{|D|(|D| - 1)}{k(k+1)}.$$

Summing all these inequalities for $k = m, m+1, \ldots, n-1$, we establish that

$$\begin{aligned}
0 \leq \max_{\substack{|T'|=m \\ T' \text{ } d\text{-ary tree}}} \gamma(D, T') - \max_{\substack{|T''|=n \\ T'' \text{ } d\text{-ary tree}}} \gamma(D, T'') &\leq |D|(|D| - 1) \sum_{k=m}^{n-1} \frac{1}{k(k+1)} \\
&= |D|(|D| - 1)\left(\frac{1}{m} - \frac{1}{n}\right).
\end{aligned}$$

Thus

$$\max_{\substack{|T'|=m\\ T' \text{ } d\text{-ary tree}}} \gamma(D,T') - \max_{\substack{|T''|=n\\ T'' \text{ } d\text{-ary tree}}} \gamma(D,T'') = \mathcal{O}\Big(\frac{n-m}{m\cdot n}\Big)$$

as $m \le n$ and $m \to \infty$. In particular,

$$\gamma(D,T_{n,1}) - \gamma(D,T_n) \le \max_{\substack{|T'|=|T_{n,1}|\\ T' \text{ } d\text{-ary tree}}} \gamma(D,T') - \gamma(D,T_n) = \mathcal{O}\Big(\frac{|T_n|-|T_{n,1}|}{|T_n|\cdot|T_{n,1}|}\Big)\,. \tag{5.3.3}$$

Using (5.3.3), formula (5.3.2) implies that

$$c(D,T_n) \le \frac{\binom{|T_{n,1}|}{|D|}}{\binom{|T_n|}{|D|}} c(D,T_n) + \mathcal{O}\Big(|T_{n,1}|^{|D|}\cdot\frac{|T_n|-|T_{n,1}|}{|T_n|\cdot|T_{n,1}|}\Big) + \mathcal{O}(\beta_n^2\cdot|T_n|^{|D|})\,.$$

Thus,

$$\Big(1-\frac{\binom{|T_{n,1}|}{|D|}}{\binom{|T_n|}{|D|}}\Big)c(D,T_n) \le \mathcal{O}\big(|T_{n,1}|^{|D|-1}\cdot\beta_n\big) + \mathcal{O}(\beta_n^2\cdot|T_n|^{|D|})$$

and using the asymptotic formula

$$\binom{|T_n|}{|D|} - \binom{|T_{n,1}|}{|D|} \sim (|T_n|-|T_{n,1}|)\frac{|T_n|^{|D|-1}}{|D|!}\,, \tag{5.3.4}$$

which holds since $|T_n| \sim |T_{n,1}|$, we derive that

$$\gamma(D,T_n) \le \mathcal{O}(|T_n|^{-1}) + \mathcal{O}(\beta_n)\,.$$

Therefore,

$$I_d(D) = \lim_{n\to\infty}\gamma(D,T_n) \le 0$$

as $\lim_{n\to\infty}\beta_n = 0$. This contradicts Proposition 3.5.0.1 in Chapter 3, which states that $I_d(D)$ is strictly positive: our second claim is proved.

Now we can assume that D has only two branches, one of which is the tree that has only one vertex. Since $|D| > 2$ by assumption, let D_2 be the second branch of D with at least two leaves. Then using Claim 1, we get

$$c(D,T_n) = c(D,T_{n,1}) + (|T_n|-|T_{n,1}|)c(D_2,T_{n,1}) + \mathcal{O}(\beta_n^2\cdot|T_n|^{|D|})\,.$$

Following the same course of reasoning used to prove Claim 2, we obtain

$$c(D, T_n) \leq \frac{\binom{|T_{n,1}|}{|D|}}{\binom{|T_n|}{|D|}} c(D, T_n) + \mathcal{O}\big(|T_n|^{|D|-1} \cdot \beta_n\big) + (|T_n| - |T_{n,1}|)c(D_2, T_{n,1}) \\ + \mathcal{O}(\beta_n^2 \cdot |T_n|^{|D|})$$

which implies that

$$\Big(1 - \frac{\binom{|T_{n,1}|}{|D|}}{\binom{|T_n|}{|D|}}\Big) c(D, T_n) \leq \mathcal{O}\big(|T_n|^{|D|-1} \cdot \beta_n\big) + (|T_n| - |T_{n,1}|)\binom{|T_{n,1}|}{|D_2|}\gamma(D_2, T_{n,1}) \\ + \mathcal{O}(\beta_n^2 \cdot |T_n|^{|D|})\,.$$

It follows from the asymptotic formula (5.3.4) that

$$\gamma(D, T_n) - \gamma(D_2, T_{n,1}) \leq \mathcal{O}(|T_n|^{-1}) + \mathcal{O}(\beta_n)\,.$$

Applying lim inf to both sides of this inequality, we get

$$I_d(D) - \limsup_{n\to\infty} \gamma(D_2, T_{n,1}) = \liminf_{n\to\infty} \big(\gamma(D, T_n) - \gamma(D_2, T_{n,1})\big) \leq 0\,.$$

Hence,

$$I_d(D) \leq \limsup_{n\to\infty} \gamma(D_2, T_{n,1}) \leq I_d(D_2)\,.$$

This completes the proof of the theorem once we invoke the second part of Lemma 5.3.0.3. □

The following corollary is a direct consequence of Theorem 5.3.0.2 combined with the first part of Lemma 5.3.0.3:

Corollary 5.3.0.4. For a d-ary tree D with branches $D_1, D_2, \ldots, D_r$ in which branches with the same number of leaves are isomorphic, we have

$$I_d(D) \leq \prod_{i=1}^{r} I_d(D_i)\,.$$

A binary tree is called a binary caterpillar if all its non-leaf vertices form a single path, beginning at the root. We denote the k-leaf binary caterpillar by F_k^2

Corollary 5.3.0.5. Let D be a d-ary tree with two branches D_1, D_2 such that $|D_1| = 1$. Then we have

$$I_d(D) \leq I_d(D_2)\,,$$

with equality for $D = F^2_{|D|}$.

Proof. It is proved in Theorem 3.2.0.2 (Chapter 3) that for every k and d, the inducibility of the k-leaf binary caterpillar F^2_k is 1 in d-ary trees. $\square$

We conjecture that the bound in Corollary 5.3.0.5 is attained only for binary caterpillars.

CONJECTURE 5.3.0.6. The inequality in Corollary 5.3.0.5 holds with equality if and only if D is a binary caterpillar.

We note that the general upper bound (Theorem 5.3.0.2) on the inducibility $I_d(D)$ can be improved considerably if one restricts to a special class of d-ary trees. For instance, when $r = d$ and the number of leaves $|D_i|$ in the branches D_i of D satisfy $\big||D_i| - |D_j|\big| \leq 1$ for all $1 \leq i, j \leq r$, we can actually calculate the supremum explicitly. Moreover, we can also show that the bound is again sharp (see the next section).

We close this section by providing a proof of Lemma 5.3.0.3:

Proof of Lemma 5.3.0.3. Let $V_{d,|D|}$ be defined by

$$V_{d,|D|} = \Big\{(i_1, i_2, \ldots, i_d) : i_1, i_2, \ldots, i_d \text{ nonnegative integers,}$$
$$i_1 + i_2 + \cdots + i_d = |D|, \text{ and } i_1, i_2, \ldots, i_d < |D|\Big\}\,.$$

Denote by $V^*_{d,|D|}$ the subset of $V_{d,|D|}$ consisting of elements $(i_1, i_2, \ldots, i_d)$ which are different from every permutation of the set

$$\{|D_1|, |D_2|, \ldots, |D_r|, \underbrace{0, 0, \ldots, 0}_{(d-r)\ 0's}\}\,.$$

Assume that branches of D that have the same number of leaves are isomorphic. Then two branches of D are isomorphic if and only if they have the same number of leaves; so we can rewrite $1-\sum_{i=1}^{d} x_i^{|D|}$ by means of the multinomial theorem as follows:

$$1-\sum_{i=1}^{d} x_i^{|D|} = \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \binom{|D|}{|D_1|,|D_2|,\ldots,|D_r|} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|} + \sum_{(i_1,i_2,\ldots,i_d)\in V^*_{d,|D|}} \binom{|D|}{i_1,i_2,\ldots,i_d} \prod_{j=1}^{d} x_j^{i_j}. \tag{5.3.5}$$

From this, we immediately deduce the inequality

$$1-\sum_{i=1}^{d} x_i^{|D|} \geq \binom{|D|}{|D_1|,|D_2|,\ldots,|D_r|} \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|}$$

showing that

$$\sup_{\substack{0\leq x_1,x_2,\ldots,x_d<1\\ x_1+x_2+\cdots+x_d=1}} \left\{ \frac{1}{1-\sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|} \right\}$$

is at most

$$\binom{|D|}{|D_1|,|D_2|,\ldots,|D_r|}^{-1}.$$

This proves the first part of the lemma. For the proof of the second part, set

$$H(D;x_1,x_2,\ldots,x_d) := \frac{\sum_{\{i_1,i_2\}\subseteq\{1,2,\ldots,d\}} \left(x_{i_1}\cdot x_{i_2}^{|D|-1} + x_{i_1}^{|D|-1}\cdot x_{i_2}\right)}{1-\sum_{i=1}^{d} x_i^{|D|}}.$$

We note that

$$\begin{aligned} H(D;x_1,x_2,\ldots,x_d) &= \frac{\left(\sum_{i=1}^{d} x_i\right)\left(\sum_{i=1}^{d} x_i^{|D|-1}\right) - \left(\sum_{i=1}^{d} x_i^{|D|}\right)}{1-\sum_{i=1}^{d} x_i^{|D|}} \\ &= \frac{x_d^{|D|-1}(1-x_d) + \sum_{i=1}^{d-1} x_i^{|D|-1}(1-x_i)}{(1-x_d)\left(\sum_{i=0}^{|D|-1} x_d^{i}\right) - \left(\sum_{i=1}^{d-1} x_i^{|D|}\right)}, \end{aligned}$$

and so

$$\lim_{\epsilon\to 0} H(D;\underbrace{\epsilon,\epsilon,\ldots,\epsilon}_{(d-1)\ \epsilon' s},1-(d-1)\epsilon) = |D|^{-1}$$

for $|D| > 2$. Hence, together with the first part of the lemma, we obtain

$$\sup_{\substack{0\le x_1,x_2,\dots,x_d<1\\ x_1+x_2+\cdots+x_d=1}} \left\{ \frac{1}{1-\sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1,i_2\}\subseteq\{1,2,\dots,d\}} \left(x_{i_1}\cdot x_{i_2}^{|D|-1} + x_{i_1}^{|D|-1}\cdot x_{i_2}\right) \right\} = |D|^{-1},$$

which completes the proof of the lemma. □

5.4 On balanced trees

In this section, we focus on a special class of topological trees which we call *balanced*.

Definition 5.4.0.1. A topological tree will be called *balanced* if the number of leaves in its branches (branch sizes) pairwise differ at most by one.

For our purposes, the tree that has only one vertex is also considered as a balanced tree. Figure 5.1 shows an example of a balanced 5-ary tree.

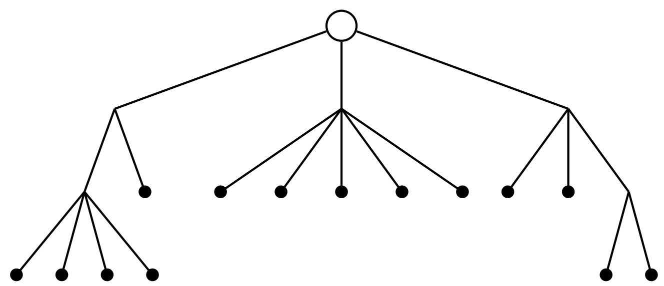


Figure 5.1: A balanced 5-ary tree.

For given positive integers $p, n \ge 2$, denote by $W_d(p,n)$ the set of all vectors $(l_1, l_2, \dots, l_d)$ of nonnegative integers for which $l_1 + l_2 + \cdots + l_d = n$, $l_1, l_2, \dots, l_d < n$ and at least $d - p$ entries equal to 0.

In what follows, we give an improved upper bound on the inducibility of every balanced d-ary tree with arbitrary root degree, and also prove, among other things, that the lower bound on $I_d(D)$ derived in Theorem 5.3.0.1 is attained for every complete d-ary tree (a special class of balanced d-ary tree).

Theorem 5.4.0.2. For a balanced d-ary tree D with branches $D_1, D_2, \ldots, D_r$, the inequality

$$I_d(D) \leq \binom{d}{r} \frac{|M(D)|}{L_d(D)} \binom{|D|}{|D_1|, |D_2|, \ldots, |D_r|} \prod_{i=1}^{r} I_d(D_i)$$

is satisfied for every d, with

$$L_d(D) = \sum_{(l_1, l_2, \ldots, l_d) \in W_d(r, |D|)} \binom{|D|}{l_1, l_2, \ldots, l_d}.$$

Furthermore, if $r = d$ and for every $i \in \{1, 2, \ldots, d\}$, the family of complete d-ary trees yields the inducibility of D_i in the limit, then the sequence of complete d-ary trees is also asymptotically maximal for the tree D, and we have

$$I_d(D) = |M(D)| \frac{\binom{|D|}{|D_1|, |D_2|, \ldots, |D_d|}}{d^{|D|} - d} \prod_{i=1}^{d} I_d(D_i).$$

Let us postpone the proof of Theorem 5.4.0.2 and provide some illustrations.

Our Theorem 5.4.0.2 can be used to yield a new approach to a result from [1]. An *even* binary tree (as defined in [1]) is any binary tree with the property that, for every internal vertex, the number of leaves in the two subtrees below it differ at most by one.

Clearly, there is only one even binary tree for any given number of leaves. We show in Figure 5.2 the even binary tree with seven leaves.

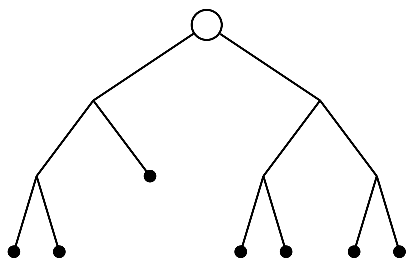


Figure 5.2: An even binary tree.

Corollary 5.4.0.3. If E_k^2 denotes the unique even binary tree with k leaves, then for every $k \geq 2$, we have

$$I_2(E_k^2) = \frac{1}{2^k - 2}\begin{cases} \binom{k}{k/2} I_2(E_{k/2}^2)^2 & \text{if } k \text{ is even}\,, \\ 2\binom{k}{\frac{k-1}{2}} I_2\left(E_{\frac{k-1}{2}}^2\right) I_2\left(E_{\frac{k+1}{2}}^2\right) & \text{otherwise}\,. \end{cases}$$

Proof. The assertion holds trivially for $k = 2$. We may then assume that it is also true for even binary trees with at most $k-1$ leaves for some $k \geq 3$ and proceed by induction on k. Consider the even binary tree with k leaves. The branches of E_k^2 are the even binary trees $E_{\lfloor k/2 \rfloor}^2$ and $E_{\lceil k/2 \rceil}^2$ by definition.

According to the induction hypothesis, the family of complete binary trees yields the inducibilities of $E_{\lfloor k/2 \rfloor}^2$ and $E_{\lceil k/2 \rceil}^2$ in the limit. Thus, by Theorem 5.4.0.2, the family of complete binary trees is also asymptotically maximal for the tree E_k^2, and the value of $I_2(E_k^2)$ is

$$|M(E_k^2)| \frac{\binom{k}{\lfloor k/2 \rfloor, \lceil k/2 \rceil}}{2^k - 2} I_2(E_{\lfloor k/2 \rfloor}^2) I_2(E_{\lceil k/2 \rceil}^2)\,.$$

Note that $|M(E_k^2)|$ is 1 if k is even, and 2 if k is odd. This completes the induction step and thus the proof of the corollary. $\square$

The next corollary provides a formula for the inducibility of a complete d-ary tree:

Corollary 5.4.0.4. For the complete d-ary tree of height h, we have

$$I_d(CD_h^d) = (d^h)! \cdot \prod_{i=0}^{h-1} \left(d^{d^{h-i}} - d\right)^{-d^i}$$

for every d and $h \geq 1$.

Proof. We fix d and demonstrate the result by induction on h. For $h = 1$, the tree CD_h^d corresponds to the d-leaf star C_d whose inducibility is shown in Theorem 3.2.0.1 (Chapter 3) to be equal to

$$\lim_{h\to\infty} \gamma(C_d, CD_h^d) = \frac{(d-1)!}{d^{d-1} - 1}\,,$$

meaning that the formula holds for $h = 1$. Assume the statement is true for every complete d-ary tree of height at most $h-1$ for some $h \geq 2$. That is,

$$I_d(CD^d_{h-1}) = \lim_{H\to\infty} \gamma(CD^d_{h-1}, CD^d_H) \, .$$

Then by Theorem 5.4.0.2, the sequence of complete d-ary trees is asymptotically maximal for the tree CD^d_h and the inducibility is given by

$$\frac{1}{d^{d^h} - d} \binom{d^h}{\underbrace{d^{h-1}, d^{h-1}, \ldots, d^{h-1}}_{d \text{ times}}} \left((d^{h-1})! \cdot \prod_{i=0}^{h-2} \left(d^{d^{h-1-i}} - d \right)^{-d^i} \right)^d$$

as $|M(CD^d_h)| = 1$. A simple manipulation reduces this latter expression to the desired identity of the corollary. $\square$

Let $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_n)$ be vectors of real numbers. Assume $a_1 \geq a_2 \geq \cdots \geq a_n$ and $b_1 \geq b_2 \geq \cdots \geq b_n$ in this order. We say that the vector A *majorises* the vector B if

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \, ,$$

and for every $k \in \{1, 2, \ldots, n-1\}$,

$$\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} b_i \, .$$

Before we get to a proof of Theorem 5.4.0.2, let us first introduce a key auxiliary lemma which provides both an upper bound on the supremum of certain functions in several variables, and the maxima of the same functions in a special case.

For given positive integers $p \geq 2$ and $n \geq 2$, set

$$V_{p,n} := \Big\{ (i_1, i_2, \ldots, i_p) : i_1, i_2, \ldots, i_p \text{ nonnegative integers,} \\ i_1 + i_2 + \cdots + i_p = n, \text{ and } i_1, i_2, \ldots, i_p < n \Big\} \, .$$

The size of $V_{p,n}$ is just $\binom{p+n-1}{n} - p$, i.e., the number of ways to choose n elements from a set of p elements, with repetitions allowed and no elements occurring n times.

Lemma 5.4.0.5. For every balanced d-ary tree D with branches $D_1, D_2, \ldots, D_r$, we have

$$\sup_{\substack{0 \leq x_1, x_2, \ldots, x_d < 1 \\ x_1 + x_2 + \cdots + x_d = 1}} \left\{ \frac{1}{1 - \sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1, i_2, \ldots, i_r\} \subseteq \{1,2,\ldots,d\}} \sum_{\pi \in M(D)} \prod_{j=1}^{d} x_{i_j}^{|D_{\pi(j)}|} \right\} \leq \binom{d}{r} \frac{|M(D)|}{L_d(D)}$$

with

$$L_d(D) = \sum_{(l_1, l_2, \ldots, l_d) \in W_d(r, |D|)} \binom{|D|}{l_1, l_2, \ldots, l_d}.$$

Furthermore, if $r = d$ then we also have

$$\max_{\substack{0 \leq x_1, x_2, \ldots, x_d < 1 \\ x_1 + x_2 + \cdots + x_d = 1}} \left\{ \frac{1}{1 - \sum_{i=1}^{d} x_i^{|D|}} \sum_{\pi \in M(D)} \prod_{j=1}^{d} x_j^{|D_{\pi(j)}|} \right\} = \frac{|M(D)|}{d^{|D|} - d}.$$

Proof. Let D be a balanced d-ary tree with r branches $D_1, D_2, \ldots, D_r$ for some $r \in \{2, 3, \ldots, d\}$. There exists a positive integer k such that each branch of D contains either k or $k+1$ leaves. So we deduce that the representatives of the equivalence classes of the equivalence relation $\sim_D$ defined on the set of all permutations of $\{1, 2, \ldots, r\}$ are exactly those permutations of $\{\underbrace{k, k, \ldots, k}_{(r-\beta)\text{ times}}, \underbrace{k+1, k+1, \ldots, k+1}_{\beta \text{ times}}\}$ induced by $\sim_D$ for some $\beta \in \{0, 1, \ldots, r-1\}$—see the discussion in Section 5.2. So we lose no generality by assuming that $|D_1| = |D_2| = \cdots = |D_{r-\beta}| = k$ and $|D_{r-\beta+1}| = |D_{r-\beta+2}| = \cdots = |D_r| = k+1$ for some $k \geq 1$ and some $\beta \in \{0, 1, \ldots, r-1\}$ (note, however, that branches with the same number of leaves can be identical).

In order to prove the lemma, we first show that every vector $(i_1, i_2, \ldots, i_r) \in V_{r, r\cdot k + \beta}$ majorises

$$(\underbrace{k+1, k+1, \ldots, k+1}_{\beta \text{ times}}, \underbrace{k, k, \ldots, k}_{(r-\beta)\text{ times}}).$$

So assume $i_1 \geq i_2 \geq \cdots \geq i_r$ in this order. Fix a positive integer $m \in \{1, 2, \ldots, r-1\}$ and suppose that

$$i_1 + i_2 + \cdots + i_m < (k+1) + (k+1) + \cdots + (k+1) + k + k + \cdots + k$$

(t copies of $k+1$ followed by $m-t$ copies of k) for some $t \in \{1, 2, \ldots, \beta\}$. Then we have

$$m \cdot i_m \leq i_1 + i_2 + \cdots + i_m < t \cdot (k+1) + (m-t) \cdot k\,,$$

which implies that $m \cdot i_m < t + m \cdot k$. Thus, $i_m \leq k$ as $t \leq m$.

On the other hand, we also have

$$\begin{aligned} i_{m+1} + i_{m+2} + \cdots + i_r &= r \cdot k + \beta - (i_1 + i_2 + \cdots + i_m) \\ &> r \cdot k + \beta - (t + m \cdot k)\,, \end{aligned}$$

which implies that

$$\begin{aligned} (r-m) \cdot i_{m+1} \geq i_{m+1} + i_{m+2} + \cdots + i_r &> r \cdot k + \beta - (t + m \cdot k) \\ &= (r-m) \cdot k + \beta - t\,. \end{aligned}$$

Thus, since $r - m > 0 \geq \beta - t$, we obtain $i_{m+1} \geq k+1$.

Altogether, we have found that $i_m \leq k < k+1 \leq i_{m+1}$, meaning that $i_m < i_{m+1}$, which is a contradiction. So we conclude that every vector $(i_1, i_2, \ldots, i_r) \in V_{r, r\cdot k+\beta}$ majorises

$$\big(\underbrace{k+1, k+1, \ldots, k+1}_{\beta \text{ times}}, \underbrace{k, k, \ldots, k}_{(r-\beta) \text{ times}}\big)\,.$$

Hence, since $W_d(p, n)$ is the set of all vectors $(l_1, l_2, \ldots, l_d)$ of nonnegative integers satisfying $l_1 + l_2 + \cdots + l_d = n$, $l_1, l_2, \ldots, l_d < n$ and at least $d - p$ entries equal to 0, we deduce that every vector $(l_1, l_2, \ldots, l_d)$ belonging to the set $W_d(r, r \cdot k + \beta)$ majorises

$$\big(\underbrace{k+1, k+1, \ldots, k+1}_{\beta \text{ times}}, \underbrace{k, k, \ldots, k}_{(r-\beta) \text{ times}}, \underbrace{0, 0, \ldots, 0}_{(d-r)\ 0's}\big)\,,$$

and thus Muirhead's inequality (see Theorem 2.3.0.1 in Chapter 2) gives

$$\begin{aligned} \sum_{\pi \in S_d} \prod_{j=1}^{d} x_{\pi(j)}^{l_j} &\geq \sum_{\pi \in S_d} \prod_{j=1}^{r} x_{\pi(j)}^{|D_j|} \\ &= \frac{r! \cdot (d-r)!}{|M(D)|} \sum_{\{i_1, i_2, \ldots i_r\} \subseteq \{1, 2, \ldots, d\}} \ \sum_{(\alpha_1, \alpha_2, \ldots, \alpha_r) \in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\alpha_j}|} \end{aligned}$$

for every such vector $(l_1, l_2, \ldots, l_d) \in W_d(r, r \cdot k + \beta)$ and every vector $(x_1, x_2, \ldots, x_d)$ of positive real numbers, where S_d denotes the set of all permutations of the indices $1, 2, \ldots, d$.

On the other hand, using this inequality together with the multinomial theorem, one obtains

$$\begin{aligned}
(x_1 + x_2 + \cdots + x_d)^{r \cdot k+\beta} &= \sum_{\substack{l_1, l_2, \ldots, l_d \geq 0 \\ l_1+l_2+\cdots+l_d = r\cdot k+\beta}} \binom{r \cdot k+\beta}{l_1, l_2, \ldots, l_d} \frac{1}{d!} \sum_{\pi \in S_d} \prod_{j=1}^{d} x_{\pi(j)}^{l_j} \\
&\geq \sum_{j=1}^{d} x_j^{r\cdot k+\beta} + \sum_{(l_1, l_2, \ldots, l_d) \in W_d(r, r\cdot k+\beta)} \binom{r \cdot k+\beta}{l_1, l_2, \ldots, l_d} \frac{1}{d!} \sum_{\pi \in S_d} \prod_{j=1}^{d} x_{\pi(j)}^{l_j} \\
&\geq \sum_{j=1}^{d} x_j^{r\cdot k+\beta} + \sum_{(l_1, l_2, \ldots, l_d) \in W_d(r, r\cdot k+\beta)} \binom{r \cdot k+\beta}{l_1, l_2, \ldots, l_d} \Bigg[\frac{1}{d!} \cdot \frac{r! \cdot (d-r)!}{|M(D)|} \\
&\qquad \cdot \sum_{\{i_1, i_2, \ldots i_r\} \subseteq \{1,2,\ldots,d\}} \; \sum_{(\alpha_1, \alpha_2, \ldots, \alpha_r) \in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\alpha_j}|} \Bigg]
\end{aligned}$$

for every vector $(x_1, x_2, \ldots, x_d)$ of nonnegative real numbers. In particular, if $(x_1, x_2, \ldots, x_d)$ is a vector of nonnegative real numbers such that $x_1 + x_2 + \cdots + x_d = 1$, then we have

$$1 - \sum_{j=1}^{d} x_j^{|D|} \geq \frac{L_d(D)}{|M(D)| \binom{d}{r}} \sum_{\{i_1, i_2, \ldots i_r\} \subseteq \{1,2,\ldots,d\}} \; \sum_{(\alpha_1, \alpha_2, \ldots, \alpha_r) \in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\alpha_j}|} \tag{5.4.1}$$

with $L_d(D)$ as defined in the lemma.

Clearly, the function

$$G_D(x_1, x_2, \ldots, x_d) := \frac{1}{1 - \sum_{i=1}^{d} x_i^{|D|}} \sum_{\{i_1, i_2, \ldots, i_r\} \subseteq \{1,2,\ldots,d\}} \; \sum_{\pi \in M(D)} \prod_{j=1}^{r} x_{i_j}^{|D_{\pi(j)}|},$$

subject to the constraint $x_1 + x_2 + \cdots + x_d = 1$, is well-defined in the region covered by the inequalities $0 \leq x_1, x_2, \ldots, x_d < 1$. Hence, we establish – by relation (5.4.1) – that

$$G_D(x_1, x_2, \ldots, x_d) \leq \frac{|M(D)|}{L_d(D)} \binom{d}{r}$$

for all $0 \leq x_1, x_2, \ldots, x_d < 1$ such that $x_1 + x_2 + \cdots + x_d = 1$, as $L_d(D)$ is never zero. This also shows that

$$\sup_{\substack{0 \leq x_1, x_2, \ldots, x_d < 1 \\ x_1 + x_2 + \cdots + x_d = 1}} G_D(x_1, x_2, \ldots, x_d) \leq \binom{d}{r} \frac{|M(D)|}{L_d(D)} .$$

Furthermore, if $r = d$, then we have

$$G_D(x_1, x_2, \ldots, x_d) = \frac{1}{1 - \sum_{i=1}^d x_i^{|D|}} \sum_{\pi \in M(D)} \prod_{j=1}^d x_j^{|D_{\pi(j)}|},$$

and a simple computation yields

$$G_D\big(\underbrace{d^{-1}, d^{-1}, \ldots, d^{-1}}_{d \text{ times}}\big) = \frac{|M(D)|}{d^{|D|} - d},$$

while the multinomial theorem gives

$$\begin{aligned} L_d(D) &= \sum_{(l_1, l_2, \ldots, l_d) \in W_d(d, |D|)} \binom{|D|}{l_1, l_2, \ldots, l_d} \\ &= \sum_{\substack{l_1 + l_2 + \cdots + l_d = |D| \\ 0 \leq l_1, l_2, \ldots, l_d < |D|}} \binom{|D|}{l_1, l_2, \ldots, l_d} = d^{|D|} - d . \end{aligned}$$

This proves that in the case where $r = d$, we have

$$\max_{\substack{0 \leq x_1, x_2, \ldots, x_d < 1 \\ x_1 + x_2 + \cdots + x_d = 1}} G_D(x_1, x_2, \ldots, x_d) = \frac{|M(D)|}{d^{|D|} - d},$$

which completes the proof of the entire lemma. □

Armed with Lemma 5.4.0.5, we can now prove Theorem 5.4.0.2.

Proof. Let D be a balanced d-ary tree with branches $D_1, D_2, \ldots, D_r$ for some $r \in \{2, 3, \ldots, d\}$. Further, set

$$c(D_j) := \frac{I_d(D_j)}{|D_j|!}, \quad \text{and} \quad c(D) := \binom{d}{r} \frac{|M(D)|}{L_d(D)} \prod_{j=1}^r c(D_j)$$

with $L_d(D)$ be as defined in the theorem. We are going to prove that there exists a nonnegative absolute constant $K(D)$ such that the inequality

$$c(D, T) \leq c(D)|T|^{|D|} + K(D)|T|^{|D|-1} \tag{5.4.2}$$

holds for every strictly d-ary tree T.

We know from the proof of Theorem 3.3.0.1 in Chapter 3 that for every d-ary tree D', the maximum of $\gamma(D',T')$ over all d-ary trees T' with $n \geq |D'|$ leaves is bounded from above by

$$I_d(D') + \frac{|D'|(-1+|D'|)}{n}\,.$$

In particular, we have

$$c(D',T) \leq \left(I_d(D') + \frac{|D'|(-1+|D'|)}{|T|}\right)\binom{|T|}{|D'|}$$

for every strictly d-ary tree T, which implies that

$$c(D',T) \leq \frac{I_d(D')}{|D'|!}\cdot|T|^{|D'|} + \frac{1}{(|D'|-2)!}\cdot|T|^{|D'|-1}$$

for every strictly d-ary tree T and d-ary tree D' such that $|D'|>2$. Taking $K(D_j)=0$ if $|D_j|\leq 2$, and $K(D_j)=1/(|D_j|-2)!$ otherwise, we see that

$$c(D_j,T) \leq c(D_j)|T|^{|D_j|} + K(D_j)|T|^{|D_j|-1} \tag{5.4.3}$$

for every $j\in\{1,2,\ldots,r\}$ as $c(D_j)=I_d(D_j)/|D_j|!$ by definition.

On the other hand, since $K(D)\geq 0$ by definition, we also see that inequality (5.4.2) holds trivially for $|T|<|D|$ as there cannot be any copies of D in T. We may then continue by induction on $|T|$.

For the induction step, consider a strictly d-ary tree T with branches $T_1,T_2,\ldots,T_d$ such that $|T|\geq|D|$. We have the following recursion – see equation (5.2.1):

$$c(D,T)=\sum_{i=1}^{d}c(D,T_i)+\sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}}\ \sum_{\pi\in M(D)}\ \prod_{j=1}^{r}c\big(D_{\pi(j)},T_{i_j}\big)\,.$$

Employing the induction hypothesis on every $c(D,T_i)$ together with (7.2.5), we get

$$\begin{aligned}
c(D,T) &\leq \sum_{i=1}^{d} \Big(c(D)|T_i|^{|D|} + K(D)|T_i|^{|D|-1} \Big) \\
&+ \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} \Big(c\big(D_{\pi(j)}\big)|T_{i_j}|^{|D_{\pi(j)}|} + K\big(D_{\pi(j)}\big)|T_{i_j}|^{|D_{\pi(j)}|-1} \Big) \\
&= c(D)\sum_{i=1}^{d}|T_i|^{|D|} + \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} \prod_{j=1}^{r} c\big(D_{\pi(j)}\big)|T_{i_j}|^{|D_{\pi(j)}|} \\
&+ K(D)\sum_{i=1}^{d}|T_i|^{|D|-1} + \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \sum_{\pi\in M(D)} A\big(d;\pi;D;T;\{i_1,i_2,\ldots,i_r\}\big)
\end{aligned} \tag{5.4.4}$$

with

$$\begin{aligned}
A\big(d;\pi;D;T;\{i_1,i_2,\ldots,i_r\}\big) = &\prod_{j=1}^{r} \Big(c\big(D_{\pi(j)}\big)|T_{i_j}|^{|D_{\pi(j)}|} + K\big(D_{\pi(j)}\big)|T_{i_j}|^{|D_{\pi(j)}|-1} \Big) \\
&- \prod_{j=1}^{r} c\big(D_{\pi(j)}\big)|T_{i_j}|^{|D_{\pi(j)}|}.
\end{aligned} \tag{5.4.5}$$

Every single term in the expansion of $A\big(d;\pi;D;T;\{i_1,i_2,\ldots,i_r\}\big)$ is of the form

$$\prod_{j=1}^{r} c\big(D_{\pi(j)}\big)^{a_j} K\big(D_{\pi(j)}\big)^{b_j} |T_{i_j}|^{a_j\cdot|D_{\pi(j)}|+b_j\cdot(|D_{\pi(j)}|-1)}$$

for some binary ($a_j, b_j \in \{0,1\}$) vector $(a_1,a_2,\ldots,a_r,b_1,b_2,\ldots,b_r)$ satisfying $a_j+b_j=1$ for all j, and $(b_1,b_2,\ldots,b_r)$ different from the null vector. Thus, $A\big(d;\pi;D;T;\{i_1,i_2,\ldots,i_r\}\big)$ contains no terms of the form

$$|T_i|^{|D|-1}\prod_{j=1}^{r} c\big(D_{\pi(j)}\big)^{a_j} K\big(D_{\pi(j)}\big)^{b_j}$$

unless $r=2$ and $|D|\leq 3$, in which case $A\big(d;\pi;D;T;\{i_1,i_2,\ldots,i_r\}\big)=0$ as $K(D_1)=K(D_2)=0$ by definition. Based on this discussion, it follows from (5.4.5) that

$$\begin{aligned}
A\big(d;\pi;D;T;\{i_1,i_2,\ldots,i_r\}\big) \leq &\left(|T|^{|D|-1} - \sum_{i=1}^{d} |T_i|^{|D|-1} \right) \\
&\cdot \left(\prod_{j=1}^{r} \Big(c\big(D_{\pi(j)}\big) + K\big(D_{\pi(j)}\big) \Big) - \prod_{j=1}^{r} c\big(D_{\pi(j)}\big) \right)
\end{aligned}$$

as all the terms of the form

$$\prod_{j=1}^{r} |T_{i_j}|^{a_j \cdot |D_{\pi(j)}| + b_j \cdot (|D_{\pi(j)}|-1)}$$

are bounded above by a term in the expansion of $\left(\sum_{i=1}^{d} |T_i|\right)^{|D|-1} = |T|^{|D|-1}$ other than one of the $|T_i|^{|D|-1}$, and none of them is $|T_i|^{|D|-1}$.

Now we can take

$$\begin{aligned} K(D) &= \sum_{\{i_1,i_2,\ldots,i_r\} \subseteq \{1,2,\ldots,d\}} \sum_{\pi \in M(D)} \left(\prod_{j=1}^{r} \left(c(D_{\pi(j)}) + K(D_{\pi(j)}) \right) - \prod_{j=1}^{r} c(D_{\pi(j)}) \right) \\ &= \binom{d}{r} \sum_{\pi \in M(D)} \left(\prod_{j=1}^{r} \left(c(D_{\pi(j)}) + K(D_{\pi(j)}) \right) - \prod_{j=1}^{r} c(D_{\pi(j)}) \right). \end{aligned}$$

Inequality (5.4.4) implies that

$$\begin{aligned} c(D,T) \leq c(D) \sum_{i=1}^{d} |T_i|^{|D|} &+ \sum_{\{i_1,i_2,\ldots,i_r\} \subseteq \{1,2,\ldots,d\}} \sum_{\pi \in M(D)} \prod_{j=1}^{r} c(D_{\pi(j)}) |T_{i_j}|^{|D_{\pi(j)}|} \\ &+ K(D)|T|^{|D|-1}. \end{aligned}$$

Finally, it remains to prove that

$$c(D) \sum_{i=1}^{d} |T_i|^{|D|} + \sum_{\{i_1,i_2,\ldots,i_r\} \subseteq \{1,2,\ldots,d\}} \sum_{\pi \in M(D)} \prod_{j=1}^{r} c(D_{\pi(j)}) |T_{i_j}|^{|D_{\pi(j)}|} \leq c(D)|T|^{|D|}$$

or equivalently,

$$\begin{aligned} \sum_{\{i_1,i_2,\ldots,i_r\} \subseteq \{1,2,\ldots,d\}} \sum_{\pi \in M(D)} \left(\prod_{j=1}^{r} c(D_{\pi(j)}) \right) \prod_{j=1}^{r} \left(\frac{|T_{i_j}|}{|T|} \right)^{|D_{\pi(j)}|} \\ \leq c(D) \left(1 - \sum_{i=1}^{d} \left(\frac{|T_i|}{|T|} \right)^{|D|} \right), \end{aligned}$$

which will then imply that $c(D,T) \leq c(D)|T|^{|D|} + K(D)|T|^{|D|-1}$ as claimed.

Since

$$\prod_{i=1}^{r} c(D_i) = \frac{c(D) L_d(D)}{|M(D)| \binom{d}{r}}$$

as defined earlier, we obtain that

$$\begin{aligned}
&\sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \quad \sum_{\pi\in M(D)} \left(\prod_{j=1}^{r} c(D_{\pi(j)})\right) \prod_{j=1}^{r} \left(\frac{|T_{i_j}|}{|T|}\right)^{|D_{\pi(j)}|} \\
&= \frac{c(D)L_d(D)}{|M(D)|\binom{d}{r}} \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \quad \sum_{\pi\in M(D)} \prod_{j=1}^{r} \left(\frac{|T_{i_j}|}{|T|}\right)^{|D_{\pi(j)}|} \\
&\leq c(D)\left(1 - \sum_{i=1}^{d} \left(\frac{|T_i|}{|T|}\right)^{|D|}\right),
\end{aligned}$$

where the inequality in the last step follows from the first part of Lemma 5.4.0.5. This completes the induction step. Consequently, taking the density of D in T and passing to the limit as $|T| \to \infty$, we obtain

$$\begin{aligned}
I_d(D) &\leq \lim_{|T|\to\infty} \frac{c(D)|T|^{|D|} + K(D)|T|^{|D|-1}}{\binom{|T|}{|D|}} = |D|! \cdot c(D) \\
&= |M(D)| \frac{|D|!}{L_d(D)} \binom{d}{r} \prod_{i=1}^{r} \frac{I_d(D_i)}{|D_i|!},
\end{aligned}$$

completing the proof of the first part of the theorem.

Now assume that $r = d$ and the sequence of complete d-ary trees is asymptotically maximal for each of the d branches of D. Then using Theorem 5.3.0.1, we get

$$I_d(D) \geq \lim_{h\to\infty} \gamma(D, CD_h^d) = \frac{|M(D)|}{d^{|D|} - d} \binom{|D|}{|D_1|, |D_2|, \ldots, |D_d|} \prod_{i=1}^{d} I_d(D_i).$$

Hence, since $L_d(D) = d^{|D|} - d$ in this case, we obtain equality:

$$I_d(D) = \lim_{h\to\infty} \gamma(D, CD_h^d).$$

This completes the proof of theorem. □

The definition of even binary trees introduced in [1] (by Czabarka, Székely and the second author of the current article) can be broadened to arbitrary d-ary trees in the following way:

Definition 5.4.0.6. Denote by $\mathcal{E}^d$ the family of d-ary trees whose elements are described recursively as follows:

- The only trees with less than d leaves in $\mathcal{E}^d$ are the stars;
- for every positive integer s, and every $\beta \in \{0, 1, 2, \ldots, d-1\}$, the tree with $n = d \cdot s + \beta$ leaves in $\mathcal{E}^d$ is obtained by attaching $d - \beta$ copies of the tree with s leaves in $\mathcal{E}^d$ and β copies of the tree with $s+1$ leaves in $\mathcal{E}^d$ to a common vertex (their respective roots are joined to a new common vertex).

The k-leaf tree in $\mathcal{E}^d$ will be denoted by E_k^d, and $\mathcal{E}^d$ will be referred to as the family of even d-ary trees. We depict in Figure 5.3 the even ternary trees with up to ten leaves.

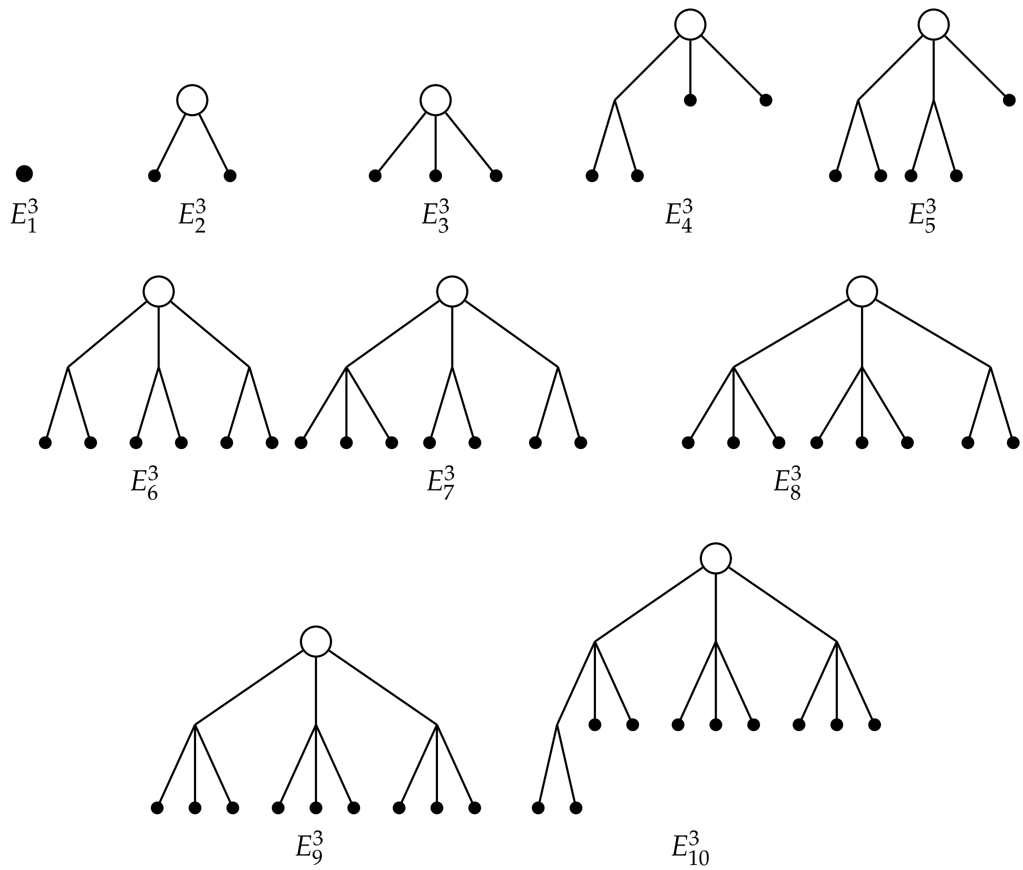


Figure 5.3: All the even ternary trees with at most ten leaves.

The next result is an asymptotic formula for the maximum density of the even d-ary tree E_r^d in strictly d-ary trees T as $|T|$ gets large. The result is (in this special case) an improvement over the formula

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D,T) = I_d(D) + \mathcal{O}\big(n^{-1/2}\big)\,,$$

which was shown in Chapter 3 to hold for general d-ary trees D.

Theorem 5.4.0.7. The inducibility of the even d-ary tree E_r^d is $r! \cdot c_r$ in d-ary trees, where c_r is defined recursively by

$$c_{d\cdot s+i} = \binom{d}{i} \frac{c_s^{d-i} \cdot c_{s+1}^i}{d^{d\cdot s+i} - d}$$

for every $s \geq 1$ and every $i \in \{0, 1, \ldots, d-1\}$, starting with $c_r = \binom{d}{r}/(d^r - d)$ for $r < d$.

Moreover, we have

$$\lim_{n\to\infty} \gamma\big(E_r^d, E_n^d\big) = I_d\big(E_r^d\big)$$

for every r, and the asymptotic formula

$$c\big(E_r^d, E_n^d\big) = c_r \cdot n^r + \mathcal{O}(n^{r-1})$$

holds for all n. In particular, we have

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma\big(E_r^d, T\big) = I_d\big(E_r^d\big) + \mathcal{O}(n^{-1})$$

for every r and all $n \equiv 1 \mod (d-1)$.

Proof. The first part of the theorem can be obtained by combining Definition 5.4.0.6 with Theorem 5.4.0.2. Indeed, the even d-ary trees E_r^d are the stars C_r for $r < d$. But we know from Theorem 3.2.0.1 in Chapter 3 that the sequence of complete d-ary trees yields $I_d(C_r)$ in the limit and

$$I_d(C_r) = \frac{d!}{(d-r)! \cdot (d^r - d)}$$

for every $r \leq d$. Therefore, it follows by induction on r that the sequence of complete d-ary trees is also asymptotically maximal for E_r^d. Assume $r \geq d$

and write $r = d \cdot s + \beta$ with β in the residue class of r modulo d. Then Theorem 5.4.0.2 gives

$$\begin{aligned} I_d(E_r^d) &= \binom{d}{\beta} \frac{(d \cdot s + \beta)!}{(s!)^{d-\beta} \cdot ((s+1)!)^{\beta} \cdot (d^{d \cdot s + \beta} - d)} \cdot I_d(E_s^d)^{d-\beta} I_d(E_{s+1}^d)^{\beta} \\ &= \binom{d}{\beta} \frac{(d \cdot s + \beta)!}{d^{d \cdot s + \beta} - d} \left(\frac{I_d(E_s^d)}{s!} \right)^{d-\beta} \left(\frac{I_d(E_{s+1}^d)}{(s+1)!} \right)^{\beta} \\ &= (d \cdot s + \beta)! \binom{d}{\beta} \frac{c_s^{d-\beta} \cdot c_{s+1}^{\beta}}{d^{d \cdot s + \beta} - d}, \end{aligned}$$

which proves the required recursive formula of the theorem.

Let us now prove that

$$\lim_{n \to \infty} \gamma(E_r^d, E_n^d) = c_r \cdot r!$$

for every r. For this purpose, let us show by means of induction with respect to r that

$$\lim_{n \to \infty} \frac{c(E_r^d, E_n^d)}{n^r} = c_r \,,$$

which already provides us with what we want. For an arbitrary but fixed nonnegative integer $\beta \leq d - 1$, we have

$$c(C_k, E_{d \cdot m + \beta}^d) = \beta \cdot c(C_k, E_{m+1}^d) + (d - \beta) \cdot c(C_k, E_m^d) + \binom{d}{k} \cdot m^k + \mathcal{O}(m^{k-1})$$

for all $m \geq 1$, and thus

$$\begin{aligned} \gamma(C_k, E_{d \cdot m + \beta}^d) = \beta \cdot \frac{\binom{m+1}{k}}{\binom{d \cdot m + \beta}{k}} \cdot \gamma(C_k, E_{m+1}^d) + (d - \beta) \cdot \frac{\binom{m}{k}}{\binom{d \cdot m + \beta}{k}} \cdot \gamma(C_k, E_m^d) \\ + \binom{d}{k} \cdot \frac{m^k}{\binom{d \cdot m + \beta}{k}} + \mathcal{O}(m^{-1}), \end{aligned}$$

when passing to the density, we deduce that

$$\begin{aligned} \liminf_{m \to \infty} \gamma(C_k, E_{d \cdot m + \beta}^d) &\geq \beta \cdot d^{-k} \cdot \liminf_{m \to \infty} \gamma(C_k, E_{m+1}^d) \\ &\quad + (d - \beta) \cdot d^{-k} \cdot \liminf_{m \to \infty} \gamma(C_k, E_m^d) + \binom{d}{k} \cdot k! \cdot d^{-k} \\ &= d^{1-k} \cdot \liminf_{m \to \infty} \gamma(C_k, E_m^d) + \binom{d}{k} \cdot k! \cdot d^{-k} \end{aligned}$$

and likewise

$$\limsup_{m\to\infty} \gamma(C_k, E^d_{d\cdot m+\beta}) \leq d^{1-k} \cdot \limsup_{m\to\infty} \gamma(C_k, E^d_m) + \binom{d}{k} \cdot k! \cdot d^{-k}.$$

Solving for the resulting equations, we obtain

$$\liminf_{n\to\infty} \gamma(C_k, E^d_n) \geq \binom{d}{k} \frac{k!}{d^k - d}, \text{ and } \limsup_{n\to\infty} \gamma(C_k, E^d_n) \leq \binom{d}{k} \frac{k!}{d^k - d},$$

from which the identity

$$\lim_{n\to\infty} \gamma(C_k, E^d_n) = \binom{d}{k} \frac{k!}{d^k - d} = I_d(C_k)$$

follows. Consequently, the statement is true for $r \leq d$, so we can focus on the induction step.

With the specialisation $T = E^d_{d\cdot m+\hat{\beta}}$ and $D = E^d_{d\cdot s+\beta}$ in the general recursion (5.2.1), we get the relation

$$\begin{aligned} c(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}}) &= (d-\beta)\cdot c(E^d_{d\cdot s+\beta}, E^d_m) + \beta \cdot c(E^d_{d\cdot s+\beta}, E^d_{m+1}) \\ &+ \sum_{(\alpha_1,\alpha_2,\ldots,\alpha_d)\in M\left(E^d_{d\cdot s+\beta}\right)} \left(\prod_{i=1}^{d-\beta} c(E^d_{\alpha_i}, E^d_m)\right)\left(\prod_{i=d-\beta+1}^{d} c(E^d_{\alpha_i}, E^d_{m+1})\right), \end{aligned}$$

which is valid for all $s, m \geq 1$ and all $\beta, \hat{\beta} \in \{0, 1, \ldots, d-1\}$. Dividing this identity through by $(d\cdot m+\hat{\beta})^{d\cdot s+\beta}$, we obtain

$$\begin{aligned} \frac{c(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}})}{(d\cdot m+\hat{\beta})^{d\cdot s+\beta}} &= (d-\beta)\cdot\left(\frac{m}{d\cdot m+\hat{\beta}}\right)^{d\cdot s+\beta} \cdot \frac{c(E^d_{d\cdot s+\beta}, E^d_m)}{m^{d\cdot s+\beta}} \\ &+ \beta\cdot\left(\frac{m+1}{d\cdot m+\hat{\beta}}\right)^{d\cdot s+\beta} \cdot \frac{c(E^d_{d\cdot s+\beta}, E^d_{m+1})}{(m+1)^{d\cdot s+\beta}} \\ &+ \sum_{(\alpha_1,\alpha_2,\ldots,\alpha_d)\in M\left(E^d_{d\cdot s+\beta}\right)} \left[\frac{m^{\sum_{i=1}^{d-\beta}|E^d_{\alpha_i}|}\cdot(m+1)^{\sum_{i=d-\beta+1}^{d}|E^d_{\alpha_i}|}}{(d\cdot m+\hat{\beta})^{d\cdot s+\beta}}\cdot\right. \\ &\left.\left(\prod_{i=1}^{d-\beta}\frac{c(E^d_{\alpha_i}, E^d_m)}{m^{|E^d_{\alpha_i}|}}\right)\left(\prod_{i=d-\beta+1}^{d}\frac{c(E^d_{\alpha_i}, E^d_{m+1})}{(m+1)^{|E^d_{\alpha_i}|}}\right)\right]. \end{aligned}$$

Applying $\liminf$ (as $m \to \infty$) to both sides of this equation, we establish that

$$\begin{aligned}
\liminf_{m\to\infty} \frac{c\left(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}}\right)}{\left(d\cdot m+\hat{\beta}\right)^{d\cdot s+\beta}} &\geq (d-\beta)\cdot d^{-(d\cdot s+\beta)} \cdot \liminf_{m\to\infty}\left(\frac{c\left(E^d_{d\cdot s+\beta}, E^d_m\right)}{m^{d\cdot s+\beta}}\right) \\
&+ \beta\cdot d^{-(d\cdot s+\beta)} \cdot \liminf_{m\to\infty}\left(\frac{c\left(E^d_{d\cdot s+\beta}, E^d_{m+1}\right)}{(m+1)^{d\cdot s+\beta}}\right) \\
&+ \sum_{(\alpha_1,\alpha_2,\ldots,\alpha_d)\in M\left(E^d_{d\cdot s+\beta}\right)} \left[d^{-(d\cdot s+\beta)} \cdot \prod_{i=1}^{d-\beta}\left(\liminf_{m\to\infty}\frac{c\left(E^d_{\alpha_i}, E^d_m\right)}{m^{|E^d_{\alpha_i}|}}\right)\cdot \right. \\
&\left. \prod_{i=d-\beta+1}^{d}\left(\liminf_{m\to\infty}\frac{c\left(E^d_{\alpha_i}, E^d_{m+1}\right)}{(m+1)^{|E^d_{\alpha_i}|}}\right)\right]
\end{aligned}$$

as we have

$$\lim_{m\to\infty} \frac{m^{\sum_{i=1}^{d-\beta}|E^d_{\alpha_i}|}\cdot(m+1)^{\sum_{i=d-\beta+1}^{d}|E^d_{\alpha_i}|}}{\left(d\cdot m+\hat{\beta}\right)^{d\cdot s+\beta}} = d^{-(d\cdot s+\beta)}$$

in view of the identity $\sum_{i=1}^{d}|E^d_{\alpha_i}| = d\cdot s+\beta$. Invoking the induction hypothesis, we arrive at

$$\begin{aligned}
\liminf_{m\to\infty} \frac{c\left(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}}\right)}{\left(d\cdot m+\hat{\beta}\right)^{d\cdot s+\beta}} &\geq \left((d-\beta)\cdot d^{-(d\cdot s+\beta)} + \beta\cdot d^{-(d\cdot s+\beta)}\right) \\
&\cdot \liminf_{m\to\infty}\left(\frac{c\left(E^d_{d\cdot s+\beta}, E^d_m\right)}{m^{d\cdot s+\beta}}\right) + \sum_{(\alpha_1,\alpha_2,\ldots,\alpha_d)\in M\left(E^d_{d\cdot s+\beta}\right)} d^{-(d\cdot s+\beta)}\left(\prod_{i=1}^{d} c_{\alpha_i}\right),
\end{aligned}$$

and this implies that

$$\left(1-d^{1-(d\cdot s+\beta)}\right)\cdot \liminf_{m\to\infty} \frac{c\left(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}}\right)}{\left(d\cdot m+\hat{\beta}\right)^{d\cdot s+\beta}} \geq d^{-(d\cdot s+\beta)}\binom{d}{\beta}\cdot c_s^{d-\beta}\cdot c_{s+1}^{\beta}.$$

Finally, we replace $\binom{d}{\beta}\cdot c_s^{d-\beta}\cdot c_{s+1}^{\beta}$ with $\left(d^{d\cdot s+\beta}-d\right)\cdot c_{d\cdot s+\beta}$ as defined in the theorem, and this gives us

$$\liminf_{m\to\infty} \frac{c\left(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}}\right)}{\left(d\cdot m+\hat{\beta}\right)^{d\cdot s+\beta}} \geq c_{d\cdot s+\beta}.$$

Similarly, taking the lim sup (as $m \to \infty$), we establish that

$$\limsup_{m\to\infty} \frac{c\big(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}}\big)}{\big(d\cdot m+\hat{\beta}\big)^{d\cdot s+\beta}} \le c_{d\cdot s+\beta}\,.$$

Therefore, we get

$$\lim_{m\to\infty} \frac{c\big(E^d_{d\cdot s+\beta}, E^d_{d\cdot m+\hat{\beta}}\big)}{\big(d\cdot m+\hat{\beta}\big)^{d\cdot s+\beta}} = c_{d\cdot s+\beta}\,,$$

which finishes the induction step. In particular, we deduce that

$$I_d\big(E^d_r\big) = \lim_{n\to\infty} \gamma\big(E^d_r, E^d_n\big) = c_r \cdot r!$$

for every r. This completes the proof of the second part of the theorem.

It remains to prove the last two assertions of the theorem. Let us first confirm that for every $r \in \{2,3,\ldots,d\}$, we have

$$c(C_r, E^d_n) \ge c_r \cdot n^r - K_r \cdot n^{r-1}$$

for all n, with $K_r = (2d)^r \cdot c_r$.

We proceed by induction on n. For $n \le d$, the tree E^d_n is the star with n leaves and so $c(C_r, E^d_n) = \binom{n}{r}$ as all the leaf-induced subtrees of a star are themselves stars. Thus, the inequality is obvious. Assume then $n > d$ and set $n = d\cdot m+\beta$, with β in the residue class of n modulo d. We have the following lower bound on the number of copies of C_r in E^d_n:

$$c\big(C_r, E^d_n\big) \ge \beta \cdot c\big(C_r, E^d_{m+1}\big) + (d-\beta)\cdot c\big(C_r, E^d_m\big) + \binom{d}{r}\cdot m^r\,,$$

and the induction hypothesis implies that

$$\begin{aligned} c\big(C_r, E^d_n\big) \ge\; & c_r\big(\beta\cdot(m+1)^r + (d-\beta)\cdot m^r\big) + \binom{d}{r}\cdot m^r \\ & - K_r\big(\beta\cdot(m+1)^{r-1} + (d-\beta)\cdot m^{r-1}\big)\,. \end{aligned}$$

Since for every positive integer $p \ge 1$, we have

$$\begin{aligned} &(d\cdot m+\beta)^p - \beta\cdot(m+1)^p - (d-\beta)\cdot m^p \\ &\qquad = (d^p - d)\cdot m^p + \sum_{l=0}^{p-1}\binom{p}{l}\Big((d\cdot m)^l\cdot\beta^{p-l} - \beta\cdot m^l\Big)\,, \end{aligned}$$

and $c_r(d^r - d) = \binom{d}{r}$ by definition, we deduce that for $r > 2$ (the case $r = 2$ is trivial, since $c(C_2, E_n^d) = \binom{n}{2}$),

$$\begin{aligned}
c_r\Big((d \cdot m + \beta)^r - \beta \cdot (m+1)^r - (d-\beta) \cdot m^r\Big) - \binom{d}{r} \cdot m^r \\
&= c_r\left(\sum_{l=0}^{r-1} \binom{r}{l} \Big(d^l \cdot \beta^{r-l} - \beta\Big) m^l\right) \\
&\leq c_r\left(\sum_{l=0}^{r-1} \binom{r}{l} d^l \cdot \beta^{r-l}\right) m^{r-1} \\
&\leq c_r(d+\beta)^r \\
&\leq K_r\left((d^{r-1} - d) \cdot m^{r-1} + \sum_{l=0}^{r-2} \binom{r-1}{l} \Big(d^l \cdot \beta^{r-1-l} - \beta\Big) m^l\right)
\end{aligned}$$

because $\beta < d$ and $K_r = (2d)^r \cdot c_r$. It follows that

$$\begin{aligned}
\beta \cdot c(C_r, E_{m+1}^d) + (d-\beta) \cdot c(C_r, E_m^d) + \binom{d}{r} \cdot m^r \\
\geq c_r(d \cdot m + \beta)^r - K_r(d \cdot m + \beta)^{r-1},
\end{aligned}$$

and this completes the induction step. Since the case $r \leq d$ has now been dealt with and the case $n < r$ is not interesting either, we can now use simultaneous induction on r and n to prove that for every r, there exists a positive absolute constant K_r (solely depending on r and d) such that

$$c\big(E_r^d, E_n^d\big) \geq c_r \cdot n^r - K_r \cdot n^{r-1}$$

for all n. In fact, we may define K_r recursively as follows:

$$K_r = (2d)^r \cdot c_r + \sum_{\pi \in M(E_r^d)} \prod_{j=1}^{d} \big(c_{|D_{\pi(j)}|} + K_{|D_{\pi(j)}|}\big), \tag{5.4.6}$$

with $D_1, D_2, \ldots, D_d$ the branches of E_r^d, starting with $K_r = (2d)^r \cdot c_r$ for $r \leq d$.

Assume that $r > d$ and $n \geq d$. So $n = d \cdot m + \beta$ with β in the residue class of n modulo d. Denote by $D_1, D_2, \ldots, D_d$ the branches of E_r^d. The recurrence relation that gives the number of copies of E_r^d in E_n^d is

$$\begin{aligned}
c\big(E_r^d, E_n^d\big) &= \beta \cdot c\big(E_r^d, E_{m+1}^d\big) + (d-\beta) \cdot c\big(E_r^d, E_m^d\big) \\
&\quad + \sum_{\pi \in M(E_r^d)} \left(\prod_{i=1}^{\beta} c\big(D_{\pi(i)}, E_{m+1}^d\big) \cdot \prod_{i=1+\beta}^{d} c\big(D_{\pi(i)}, E_m^d\big) \right) \\
&\geq \beta \cdot c\big(E_r^d, E_{m+1}^d\big) + (d-\beta) \cdot c\big(E_r^d, E_m^d\big) + \sum_{\pi \in M(E_r^d)} \prod_{i=1}^{d} c\big(D_{\pi(i)}, E_m^d\big).
\end{aligned}$$

Using the induction hypothesis, we obtain

$$\begin{aligned}
&c\big(E_r^d, E_n^d\big) \geq \beta \cdot \big(c_r \cdot (m+1)^r - K_r \cdot (m+1)^{r-1}\big) + (d-\beta) \cdot \big(c_r \cdot m^r - K_r \cdot m^{r-1}\big) \\
&+ \sum_{\pi \in M(E_r^d)} \prod_{i=1}^{d} \big(c_{|D_{\pi(i)}|} \cdot m^{|D_{\pi(i)}|} - K_{|D_{\pi(i)}|} \cdot m^{|D_{\pi(i)}|-1}\big) \\
&= \beta \cdot \big(c_r \cdot (m+1)^r - K_r \cdot (m+1)^{r-1}\big) + (d-\beta) \cdot \big(c_r \cdot m^r - K_r \cdot m^{r-1}\big) \\
&+ \sum_{\pi \in M(E_r^d)} \Bigg[\prod_{i=1}^{d} c_{|D_{\pi(i)}|} \cdot m^{|D_{\pi(i)}|} + \sum_{l=0}^{d-1} \sum_{\{i_1,\dots,i_l\} \subseteq \{1,2,\dots,d\}} \prod_{j=1}^{l} c_{|D_{\pi(i_j)}|} \cdot m^{|D_{\pi(i_j)}|} \\
&\qquad\qquad \cdot (-1)^{d-l} \cdot \prod_{\substack{i=1 \\ i \notin \{i_1,\dots,i_l\}}}^{d} K_{|D_{\pi(i)}|} \cdot m^{|D_{\pi(i)}|-1} \Bigg],
\end{aligned}$$

where the empty product is treated as 1. By neglecting the terms for which

$d-l$ is even, and replacing $|M(E_r^d)| \cdot \prod_{j=1}^{d} c_{|D_j|}$ with $c_r \cdot (d^r - d)$, we obtain

$$
\begin{aligned}
c\big(E_r^d, E_n^d\big) &\geq c_r\Big(\beta \cdot (m+1)^r + (d-\beta)\cdot m^r + (d^r - d)\cdot m^r\Big) \\
&- K_r\Big(\beta \cdot (m+1)^{r-1} + (d-\beta)\cdot m^{r-1}\Big) \\
&- \sum_{\pi \in M(E_r^d)} \Bigg[\sum_{\substack{l=0 \\ d-l \text{ odd}}}^{d-1} \sum_{\{i_1,\ldots,i_l\}\subseteq\{1,2,\ldots,d\}} \prod_{j=1}^{l} c_{|D_{\pi(i_j)}|} \cdot m^{|D_{\pi(i_j)}|} \\
&\qquad \cdot \prod_{\substack{i=1 \\ i\notin\{i_1,\ldots,i_l\}}}^{d} K_{|D_{\pi(i)}|} \cdot m^{|D_{\pi(i)}|-1} \Bigg] \\
&\geq c_r\left((d\cdot m+\beta)^r - \sum_{p=0}^{r-1} \binom{r}{p} (d^p \cdot \beta^{r-p} - \beta) m^p \right) \\
&- K_r\left((d\cdot m+\beta)^{r-1} - (d^{r-1} - d)\cdot m^{r-1} - \sum_{p=0}^{r-2} \binom{r-1}{p} (d^p \cdot \beta^{r-1-p} - \beta) m^p \right) \\
&- \sum_{\pi \in M(E_r^d)} \left(\sum_{l=0}^{d-1} \sum_{\{i_1,\ldots,i_l\}\subseteq\{1,2,\ldots,d\}} \prod_{j=1}^{l} c_{|D_{\pi(i_j)}|} \cdot \prod_{\substack{i=1 \\ i\notin\{i_1,\ldots,i_l\}}}^{d} K_{|D_{\pi(i)}|} \right) m^{r-d+l} \\
&\geq c_r \cdot (d\cdot m+\beta)^r - K_r \cdot (d\cdot m+\beta)^{r-1} - c_r \sum_{p=0}^{r-1} \binom{r}{p} (d^p \cdot \beta^{r-p} - \beta) m^p \\
&+ K_r\left((d^{r-1} - d)\cdot m^{r-1} + \sum_{p=0}^{r-2} \binom{r-1}{p} (d^p \cdot \beta^{r-1-p} - \beta) m^p \right) \\
&- \sum_{\pi \in M(E_r^d)} \left(\sum_{l=0}^{d-1} \sum_{\{i_1,\ldots,i_l\}\subseteq\{1,2,\ldots,d\}} \prod_{j=1}^{l} c_{|D_{\pi(i_j)}|} \cdot \prod_{\substack{i=1 \\ i\notin\{i_1,\ldots,i_l\}}}^{d} K_{|D_{\pi(i)}|} \right) m^{r-1}.
\end{aligned}
$$

Hence, it follows from (5.4.6) that

$$c\big(E_r^d, E_n^d\big) = c\big(E_r^d, E_{d\cdot m+\beta}^d\big) \geq c_r \cdot (d\cdot m+\beta)^r - K_r \cdot (d\cdot m+\beta)^{r-1}$$

for all $m \geq 1$ as

$$K_r \geq (2d)^r \cdot c_r + \sum_{\pi \in M(E_r^d)} \sum_{l=0}^{d-1} \sum_{\{i_1,\ldots,i_l\}\subseteq\{1,2,\ldots,d\}} \prod_{j=1}^{l} c_{|D_{\pi(i_j)}|} \cdot \prod_{\substack{i=1 \\ i\notin\{i_1,\ldots,i_l\}}}^{d} K_{|D_{\pi(i)}|}.$$

This completes the induction step.

Therefore, in view of inequality (5.4.2), which is established in the proof of Theorem 5.4.0.2, we deduce that

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma\big(E_r^d, T\big) = c_r \cdot r! + \mathcal{O}(n^{-1}),$$

which is indeed the desired result. This completes the proof of the theorem. □

The (ternary) inducibilities of the first few even ternary trees are indicated in Table 5.1.

Table 5.1: Some values of $I_3(E_k^3)$.

k	1	2	3	4	5	6	7	8	9	10	11	12
$I_3(E_k^3)$	1	1	$\frac{1}{4}$	$\frac{6}{13}$	$\frac{3}{8}$	$\frac{15}{121}$	$\frac{15}{208}$	$\frac{35}{2186}$	$\frac{7}{5248}$	$\frac{1575}{255886}$	$\frac{4725}{453596}$	$\frac{1247400}{194594881}$

We seem to possess good enough evidence to believe that the even d-ary tree E_n^d has the greatest number of copies of the tree E_r^d over all n-leaf d-ary trees. Let us put it on a more formal footing:

CONJECTURE 5.4.0.8. Let $d \geq 2$ be a fixed positive integer. For every positive integer r, we have

$$\max_{\substack{|T|=n \\ T \ d\text{-ary tree}}} c(E_r^d, T) = c(E_r^d, E_n^d)$$

for every n.

5.5 Further bounds under restriction

Besides the even d-ary trees E_r^d for which $I_d(E_r^d)$ is precisely the upper bound given in Theorem 5.4.0.2, we also remark that in certain cases, there is a lower bound that asymptotically matches the upper bound on the inducibility of D given in Theorem 5.4.0.2, as $|D|$ gets large. Our next theorem presents a result that supports this observation.

Theorem 5.5.0.1. Let $d \geq 2$ be a fixed positive integer and D a balanced d-ary tree with branches $D_1, D_2, \ldots, D_d$. Assume that each of the branches of D is a binary caterpillar. Then for $|D| \geq 4$, we have

$$\frac{|M(D)|}{d^{|D|}}\binom{|D|}{|D_1|, |D_2|, \ldots, |D_d|} \leq I_d(D) \leq \frac{|M(D)|}{d^{|D|} - d}\binom{|D|}{|D_1|, |D_2|, \ldots, |D_d|}.$$

Proof. For a fixed $d \geq 2$, denote by $X^d_{d \cdot s+\beta}$ the d-ary tree whose branches are $d-\beta$ binary caterpillars F^2_s and β binary caterpillars F^2_{s+1}, where $s \geq 1$ is any positive integer and β any nonnegative integer in $\{0, 1, \ldots, d-1\}$ (see Figure 5.4 for an illustration).

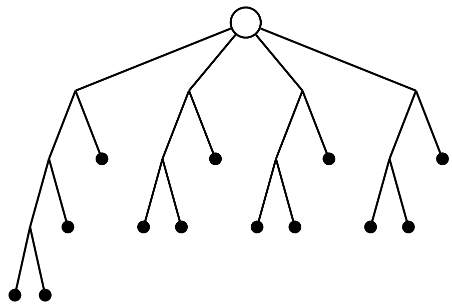


Figure 5.4: The 4-ary tree X^4_{13} described in the proof of Theorem 5.5.0.1.

Set $a_1 = a_2 = \cdots = a_{d-\beta} = s$ and $a_{d-\beta+1} = a_{d-\beta+2} = \cdots = a_d = s+1$. For $|D| \geq 4$, let us prove that for fixed $s \geq 1$ and $\beta \in \{0, 1, \ldots, d-1\}$, the identity

$$\lim_{n \to \infty} \frac{c(X^d_{d \cdot s+\beta}, X^d_n)}{\binom{n}{d \cdot s+\beta}} = \frac{\binom{d}{\beta}}{(s!)^{d-\beta} \cdot \big((s+1)!\big)^{\beta}} \cdot \frac{(d \cdot s+\beta)!}{d^{d \cdot s+\beta}}$$

holds. We begin by giving the recursion that counts the number of copies of $X^d_{d \cdot s+\beta}$ in X^d_n. Set $n = d \cdot m + \hat{\beta}$ where $\hat{\beta}$ is in the residue class of n modulo d. So we have

$$c\big(X^d_{d \cdot s+\beta}, X^d_{d \cdot m+\hat{\beta}}\big) = \sum_{(\alpha_1, \alpha_2, \ldots, \alpha_d) \in M\left(X^d_{d \cdot s+\beta}\right)} \left(\prod_{i=1}^{d-\beta} c(F^2_{a_{\alpha_i}}, F^2_m)\right)\left(\prod_{i=d-\beta+1}^{d} c(F^2_{a_{\alpha_i}}, F^2_{m+1})\right)$$

by virtue of equation (5.2.1), as evidently there cannot be any copy of $X^d_{d\cdot s+\beta}$ in F^2_m or F^2_{m+1} (all the leaf-induced subtrees of binary caterpillars are themselves binary caterpillars).

Passing to the density, we obtain

$$\gamma\big(X^d_{d\cdot s+\beta}, X^d_{d\cdot m+\hat{\beta}}\big) = \sum_{(\alpha_1,\alpha_2,\ldots,\alpha_d)\in M\left(X^d_{d\cdot s+\beta}\right)} \frac{\left(\prod_{i=1}^{d-\beta}\binom{m}{a_{\alpha_i}}\right)\left(\prod_{i=d-\beta+1}^{d}\binom{m+1}{a_{\alpha_i}}\right)}{\binom{d\cdot m+\hat{\beta}}{d\cdot s+\beta}}$$

as soon as $m \geq s+1$. Letting $m \to \infty$ and taking the limit, we arrive at

$$\lim_{m\to\infty} \gamma\big(X^d_{d\cdot s+\beta}, X^d_{d\cdot m+\hat{\beta}}\big) =$$
$$d^{-(d\cdot s+\beta)}\cdot(d\cdot s+\beta)!\cdot \sum_{(\alpha_1,\alpha_2,\ldots,\alpha_d)\in M\left(X^d_{d\cdot s+\beta}\right)} \left(\prod_{i=1}^{d} a_{\alpha_i}!\right)^{-1}.$$

Hence, we establish that

$$\lim_{m\to\infty} \gamma\big(X^d_{d\cdot s+\beta}, X^d_{d\cdot m+\hat{\beta}}\big) = \frac{|M\big(X^d_{d\cdot s+\beta}\big)|}{(s!)^{d-\beta}\cdot\big((s+1)!\big)^{\beta}}\cdot\frac{(d\cdot s+\beta)!}{d^{d\cdot s+\beta}},$$

and the first part of the theorem is proved.

The upper bound on $I_d\big(X^d_{d\cdot s+\beta}\big)$ is a consequence of Theorem 5.4.0.2 as binary caterpillars have inducibility 1 for every d—see Theorem 3.2.0.2. This completes the proof of the theorem. $\square$

Observe that for $|D| \leq 2\cdot d$, the tree $X^d_{|D|}$ is an even d-ary tree, and the upper bound on $I_d\big(X^d_{|D|}\big)$ is its precise inducibility (Theorem 5.4.0.7). However, for $|D| > 2\cdot d$, the exact inducibility of the tree $X^d_{|D|}$ is not known (unless $d = 2$ and $|D| \leq 6$). We have just established (Theorem 5.5.0.1) upper and lower bounds on the inducibility $I_d(X^d_{|D|})$, but the ratio between these two bounds approaches 1 at the rate of $d^{-|D|}$ when $|D|$ gets large. We conclude that the bounds on $I_d(X^d_{d\cdot s+\beta})$ are quite accurate for larger values of $|D|$.

In fact, we have got more: Theorem 5.5.0.1 can be refined and adapted to a more general situation.

Theorem 5.5.0.2. Let $d \geq 2$ be an arbitrary but fixed positive integer and D a d-ary tree. Assume that D has d branches all of which are isomorphic to the same d-ary tree, say D'. Then we have

$$\frac{|D|!}{d^{|D|}}\left(\frac{I_d(D')}{|D'|!}\right)^d \leq I_d(D) \leq \frac{|D|!}{d^{|D|}-d}\left(\frac{I_d(D')}{|D'|!}\right)^d .$$

Proof. The lower bound is a special case of Theorem 3.5.0.2 in Chapter 3, while the upper bound is a consequence of Theorem 5.4.0.2. □

We conclude with lower bounds, restricting ourselves to trees with only two branches.

Theorem 5.5.0.3. Let D be a d-ary tree with two branches D_1 and D_2. Then the following inequalities hold:

1. If D_1 and D_2 have the same number of leaves:

$$I_d(D) \geq 2^{-|D|}\binom{|D|}{|D|/2} I_d(D_1) I_d(D_2) ;$$

2. If D_1 and D_2 have a different number of leaves (assume $|D_1| < |D_2|$):

$$I_d(D) \geq |D|^{-|D|} \cdot |D_1|^{|D_1|} \cdot |D_2|^{|D_2|}\binom{|D|}{|D_1|} I_d(D_1) I_d(D_2) .$$

Proof. By (5.2.1), the density $\gamma(D,T)$ of D in a d-ary tree T with two branches T_1 and T_2 is at least

$$\frac{1}{\binom{|T|}{|D|}} \sum_{\pi \in M(D)} \binom{|T_1|}{|D_{\pi(1)}|}\gamma(D_{\pi(1)},T_1)\binom{|T_2|}{|D_{\pi(2)}|}\gamma(D_{\pi(2)},T_2)$$

for $|T| \geq |D|$, where we only consider copies of D in which its branches D_1, D_2 are induced by subsets of leaves of T_1, T_2.

- Suppose that $|D_1| = |D_2|$. If D_1 and D_2 are isomorphic then consider a sequence $T_1^1, T_1^2, T_1^3 \ldots$ of d-ary trees that is asymptotically maximal for D_1 (and thus D_2). Call T^n the d-ary tree with two branches, each of which is isomorphic to T_1^n. So we have

$$\gamma(D,T^n) \geq \frac{1}{\binom{2|T_1^n|}{|D|}} \binom{|T_1^n|}{|D_1|}^2 \gamma(D_1,T_1^n)^2\,,$$

(as $|M(D)| = 1$), and this implies that

$$\limsup_{n\to\infty} \gamma(D,T^n) \geq \frac{|D|!}{|D_1|!^2 \cdot 2^{|D|}} \Big(\limsup_{n\to\infty} \gamma(D_1,T_1^n)\Big)^2\,.$$

In particular, one obtains

$$I_d(D) \geq 2^{-|D|} \binom{|D|}{|D|/2} I_d(D_1)^2\,.$$

If D_1 and D_2 are not isomorphic, then consider a sequence $T_1^1, T_1^2, T_1^3 \ldots$ of d-ary trees that is asymptotically maximal for D_1 and another sequence $T_2^1, T_2^2, T_2^3 \ldots$ of d-ary trees that is asymptotically maximal for D_2. Since one can always assume that $|T_1^n| = n = |T_2^n|$, we obtain

$$\gamma(D,T^n) \geq \frac{1}{\binom{|T^n|}{|D|}} \binom{|T_1^n|}{|D|/2} \gamma(D_1,T_1^n) \binom{|T_2^n|}{|D|/2} \gamma(D_2,T_2^n)$$

where T^n is the d-ary tree with two branches, one is isomorphic to T_1^n and the other one isomorphic to T_2^n. Passing to the limit as $n \to \infty$, we deduce that

$$I_d(D) \geq \limsup_{n\to\infty} \gamma(D,T^n) \geq \frac{|D|!}{(|D|/2)!^2 \cdot 2^{|D|}} I_d(D_1) I_d(D_2)\,.$$

- Now suppose that D_1 and D_2 have different number of leaves. One can then assume that $|D_1| < |D_2|$. Choose a sequence $T_1^1, T_1^2, T_1^3 \ldots$ of d-ary trees asymptotically maximal for D_1 and another sequence $T_2^1, T_2^2, T_2^3 \ldots$ of d-ary trees asymptotically maximal for D_2 in such a way that $|T_1^n|/|T_2^n| = |D_1|/|D_2|$ for all n (which is always possible!). Let T^n be the d-ary tree with two branches, one is isomorphic to T_1^n and the other one isomorphic to T_2^n. Thus, we have

$$|T_1^n| = \frac{|D_1|}{|D|}|T^n|, \text{ and } |T_2^n| = \frac{|D_2|}{|D|}|T^n|\,.$$

Set $\alpha = |D_1|/|D|$. It follows that

$$\begin{aligned}\gamma(D, T^n) &\geq \frac{1}{\binom{|T^n|}{|D|}} \binom{|T_1^n|}{|D_1|} \gamma(D_1, T_1^n) \binom{|T_2^n|}{|D_2|} \gamma(D_2, T_2^n)\,, \\ &= \frac{1}{\binom{|T^n|}{|D|}} \binom{\alpha \cdot |T^n|}{|D_1|} \gamma(D_1, T_1^n) \binom{(1-\alpha)|T^n|}{|D_2|} \gamma(D_2, T_2^n)\,.\end{aligned}$$

and this gives us

$$\limsup_{n\to\infty} \gamma(D, T^n) \geq \alpha^{|D_1|}(1-\alpha)^{|D_2|} \binom{|D|}{|D_1|} I_d(D_1) I_d(D_2)\,.$$

In particular, we obtain

$$I_d(D) \geq |D|^{-|D|} \cdot |D_1|^{|D_1|} \cdot |D_2|^{|D_2|} \binom{|D|}{|D_1|} I_d(D_1) I_d(D_2)\,.$$

This completes the proof of the theorem. $\square$

Let us remark that one cannot improve on the lower bounds using only the argument in the proof of Theorem 5.5.0.3. This is because for any fixed positive integers k and l, the function $f(x) = x^k(1-x)^l$ on the interval $(0, 1)$ has its unique maximum at $k/(l+k)$:

$$x^k(1-x)^l \leq \frac{k^k \cdot l^l}{(l+k)^{l+k}}$$

for all $0 < x < 1$. Indeed, the first derivative of f is given by

$$f'(x) = x^{k-1}(1-x)^l(k - l \cdot x/(1-x))$$

showing that $f'(x) \geq 0$ if and only if $x \leq k/(l+k)$; in particular,

$$f(x) \leq f\Big(\frac{k}{l+k}\Big) = \frac{k^k \cdot l^l}{(l+k)^{l+k}}$$

for all $0 < x < 1$.

Remark 5.5.0.4. One derives from the proof of Theorem 5.5.0.3 that if D_1 and D_2 are not isomorphic but have a common sequence of trees that is asymptotically maximal for both, then the inducibility of D in d-ary trees is at least

$$2^{1-|D|} \binom{|D|}{|D_1|} I_d(D_1) I_d(D_2)\,.$$

Chapter 6

On the inducibility of small trees

The quantity that captures the asymptotic value of the maximum number of appearances of a given topological tree (a rooted tree with no vertices of outdegree 1) S with k leaves in an arbitrary tree with sufficiently large number of leaves is called the inducibility of S. Its precise value is known only for some specific families of trees, most of them exhibiting a symmetrical configuration. In an attempt to answer a recent question posed by Czabarka, Székely, and Wagner, we provide bounds for the inducibility $J(A_5)$ of the 5-leaf binary tree A_5 whose branches are a single leaf and the complete binary tree of height 2. The aforementioned authors indicated that $J(A_5)$ is 'close' to $1/4$. We can show that $0.24707 \leq J(A_5) \leq 0.24745$. Furthermore, we also consider the problem of determining the inducibility of the tree Q_4 (having the stars with one and three leaves as its two branches), which is the only tree among 4-leaf topological trees for which the inducibility is unknown.

The results in this chapter will appear as the following paper [38]: *On the inducibility of small trees. A. V. Dossou-Olory, and S. Wagner. To be submitted*.

6.1 Introduction and previous results

The study of graph inducibility was brought forward in 1975 by Pippenger and Golumbic, who investigated the maximum frequency of k-vertex simple graphs occurring as subgraphs within a graph whose number of vertices approaches infinity – see [2] for details and first results on the inducibility of graphs. To this day, there is substantial activity regarding this concept. Bubeck and Linial [29] defined the inducibility of a tree S with k vertices as the maximum proportion of S as a subtree among all k-vertex subtrees of a tree whose number of vertices tends to infinity. We also mention that Sperfeld [18] extended the concept of inducibility to monodirected graphs, and also gave bounds (using Razborov's flag algebra method) for some graphs with at most four vertices. In previous chapters, we were considering various inducibilities associated with rooted trees.

For any of the aforementioned notions of inducibility, can the exact inducibility of trees (graphs) with a moderate size be always determined explicitly? The answer to this question turns out to be either undecidable or negative in general in the original context of simple graphs [16; 7; 18; 29; 4]. The concept of inducibility of a tree with k leaves is still new and the precise value of the inducibility is known only for a few classes of trees, most of them exhibiting a symmetrical configuration. The recent paper [1] raised some questions on the inducibility of binary trees, one of which is discussed and approximately solved within this part of the thesis. The present chapter also covers a related problem concerning the inducibility of a ternary tree with four leaves.

The inducibility of trees with k leaves is a newly proposed quantity. The inducibility of a topological tree S (as defined and studied in Chapter 4) is its maximum density as a leaf-induced subtree of T as the size of T tends to infinity:

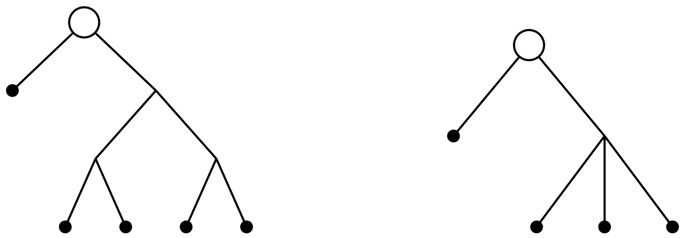
$$J(S) := \limsup_{\substack{|T|\to\infty \\ T \text{ topological tree}}} \gamma(S,T) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ topological tree}}} \gamma(S,T)\,.$$

The limit is known to exist (see Theorem 4.2.0.1 in Chapter 4). Similarly, when the underlying set over which the supremum is taken is restricted to d-ary trees, we define

$$I_d(D) := \limsup_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(D,T) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T)$$

to be the inducibility of a d-ary tree D in d-ary trees (where the limit is also known to exist—Theorem 3.3.0.1 in Chapter 3). The subscript d is used to emphasize the fact that we are taking the maximum over the set of all d-ary trees.

While in the past many results on the inducibility were obtained for graphs, this is not yet the case for trees and many challenging questions remain. The problem of computing the inducibility of a tree appears to be quite difficult even for trees with a small number of leaves—already the inducibilities of some trees with only four or five leaves are not known. Among 5-leaf binary trees, the tree A_5 (Figure 6.1) is the only one for which the inducibility has not been determined yet. Also, the inducibility of the 4-leaf ternary tree Q_4 shown in Figure 6.1 is unknown.



The binary tree A_5. The ternary tree Q_4.

Figure 6.1: The topological trees A_5 and Q_4.

In previous Chapters 3 to 5, various lower bounds were given on the inducibility of topological trees and thus the inducibilities of Q_4 and A_5. In this chapter, we shall propose constructions that yield improved lower bounds on the inducibility of the two trees Q_4 and A_5. Moreover, using a computer search, we shall be able to bound both the inducibility of A_5 in topological trees and the inducibility of Q_4 in ternary trees from above.

The inducibility of some families of topological trees is known precisely. As such, we have stars, binary caterpillars (see Chapter 3), complete d-ary trees and more generally, so-called even d-ary trees (see Chapter 5). We already know the inducibility of all topological trees with at most three leaves: each of them has inducibility 1, except for the star with three leaves, which has inducibility $(d-2)/(d+1)$ in d-ary trees. There are only five different topological trees with four leaves (see Figure 6.2), and the precise inducibility of four of them is at least partially (for some values of d)

known:

$$
\begin{aligned}
J(CD_2^2) &= I_d(CD_2^2) = \frac{3}{7} \text{ for all } d \text{ (see Chapter 5 and [1])},\\
J(F_4^2) &= I_d(F_4^2) = 1 \text{ for all } d \text{ (see Chapter 3 and [1])},\\
J(C_4) &= 1 \text{ (see Chapter 3)},\\
I_d(C_4) &= \frac{6-5d+d^2}{1+d+d^2} \text{ for all } d \text{ (see Chapter 3)},\\
I_3(E_4^3) &= \frac{6}{13} \text{ (see Chapter 5)},\\
I_d(E_4^3) &= \text{unknown for } d > 3\,.
\end{aligned}
$$

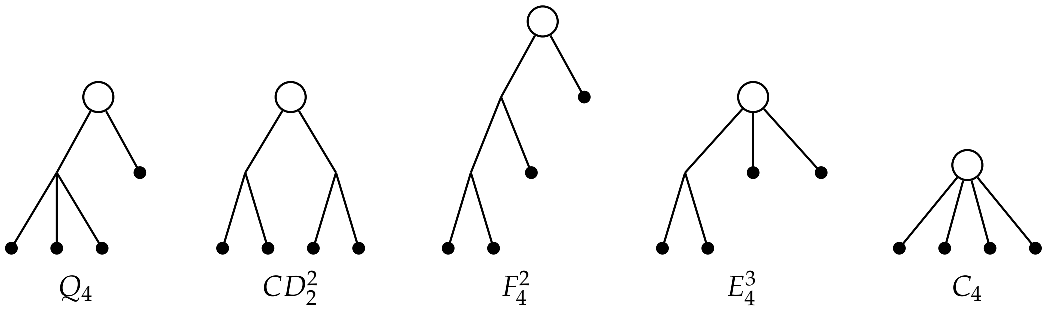


Figure 6.2: All the topological trees with four leaves.

When considering binary trees, we notice that there are only three isomorphism types of 5-leaf trees – see Figure 6.3 – and the inducibility of two of them has been determined:

$$
\begin{aligned}
J(E_5^2) &= I_d(E_5^2) = \frac{2}{3} \text{ (see Chapter 5 and [1])},\\
J(F_5^2) &= I_d(F_5^2) = 1 \text{ (see Chapter 3 and [1])}
\end{aligned}
$$

for all d. The inducibility of the binary tree A_5 is of particular interest to us, since it is the smallest binary tree for which the inducibility is not known explicitly. In [1], Czabarka, Székely, and Wagner considered the problem of computing the inducibility of the tree A_5 in binary trees, and mentioned that $I_2(A_5)$ appears to be close to $1/4$. This observation came from a computer experiment, but no explicit sequence of binary trees that would yield a value close to 0.25 in the limit was given. Here we provide a construction which yields the value 0.24707... as a lower bound. We also

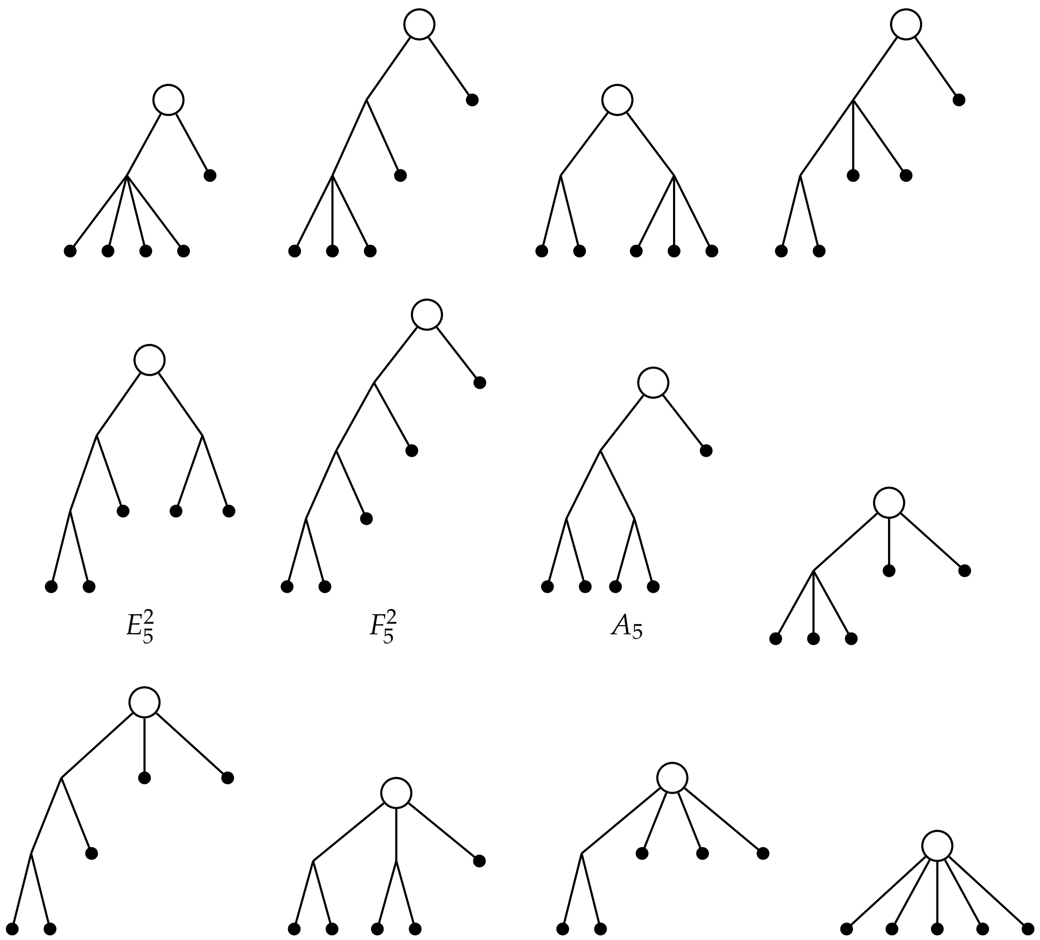


Figure 6.3: All the 5-leaf topological trees.

describe how to perform an efficient computer search and obtain 0.24745 as an upper bound on $I_2(A_5)$.

In the second part of this chapter, we consider the problem of finding the inducibility of the ternary tree Q_4 in ternary trees. Specifically, we prove that $0.1418\ldots \leq I_3(Q_4) \leq 0.1435\ldots$. These two trees that we focus on exhibit a non-symmetrical configuration, which makes the computation of their inducibilities harder. For the binary tree A_5, we are tempted to conjecture that our candidate is an optimal sequence of binary trees giving the explicit value of $I_2(A_5)$ in the limit, which we obtain as a function of the global maximum of a certain three-variable polynomial over a specific domain.

6.2 Statement of results

Chapter 4 covers, among other things, the relationship between the degree-restricted inducibility $I_d(S)$ in d-ary trees and the general inducibility $J(S)$ in topological trees at large. It was proved in Chapter 4 that

$$J(S) = \lim_{d\to\infty} I_d(S) .$$

A d-ary tree will is called a strictly d-ary tree if each of its internal vertices has exactly d children. By Theorem 3.3.0.3 in Chapter 3, we also know that the underlying set over which the maximum density in d-ary trees is taken can be reduced to strictly d-ary trees, that is

$$I_d(S) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(S,T) .$$

In [1], the authors formulated some questions and conjectures on the inducibility in binary trees, one of which was solved recently in [34] (see Chapter 3). Among the questions posed, one of them asks for the inducibility of the 5-leaf binary tree A_5 (see Figure 6.1). As mentioned in the introduction, this problem appears to be quite hard and finding a sequence of binary trees that yields $I_2(A_5)$ in the limit also appears to be a difficult task. The authors of [1] further mentioned that $I_2(A_5)$ is close to $1/4$, which will be made more precise here with the following result:

Theorem 6.2.0.1. For the binary tree A_5, we have

$$0.2470715 \leq J(A_5) = I_2(A_5) \leq 0.24745 .$$

As part of the ingredients needed to prove this result, let us define a new class of binary trees (which is already considered in recent papers [1; 37]–Chapter 5). The even binary tree E_n^2 with n leaves is obtained recursively as follows: E_1^2 is the tree with only one vertex; for $n > 1$, the branches of E_n^2 are the even binary trees $E^2_{\lfloor n/2 \rfloor}$ and $E^2_{\lceil n/2 \rceil}$. An example of an even binary tree can be found in Figure 6.4.

We shall prove the upper bound in Theorem 6.2.0.1 by means of an algorithmic approach. For the lower bound, we shall make use of the binary

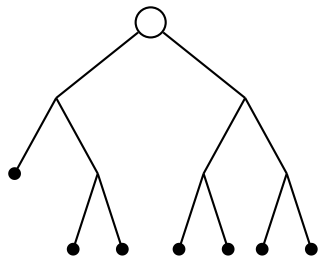


Figure 6.4: The even binary tree E_7^2 with seven leaves.

tree $S(n_1, n_2, n_3, n_4)$ whose rough picture is shown in Figure 6.5, where each triangle represents an even binary tree. More specifically, to obtain the tree $S(n_1, n_2, n_3, n_4)$, we take the 4-leaf binary tree whose internal vertices form a path beginning at the root (the square vertex on top in Figure 6.5), and identify the four leaves with the even binary trees whose number of leaves is n_1, n_2, n_3, n_4, respectively in this order (starting with the top leaf attached to the root).

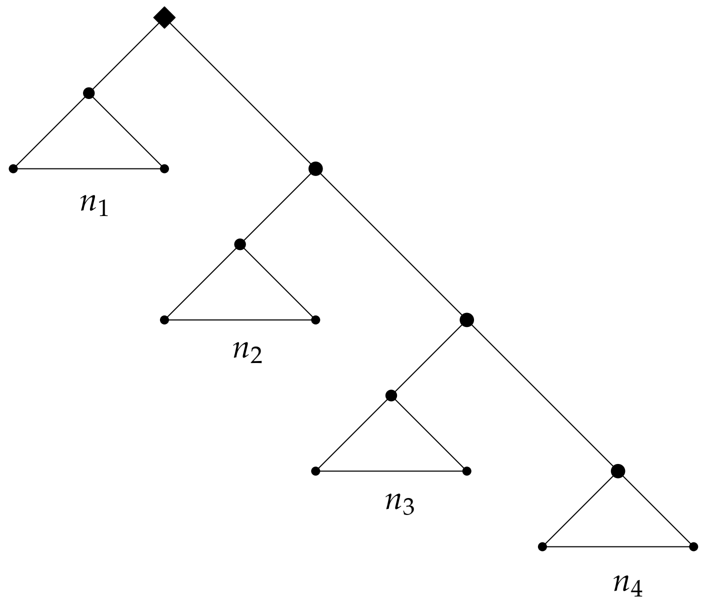


Figure 6.5: The binary tree $S(n_1, n_2, n_3, n_4)$ described for Theorem 6.2.0.1.

As a next step, we set up a formula for the number of copies of A_5 in $S(n_1, n_2, n_3, n_4)$; this formula is used together with a result on even binary trees from Chapter 5 to derive an asymptotic formula for $c(A_5, S(n_1, n_2, n_3, n_4))$ as $n = n_1 + n_2 + n_3 + n_4 \to \infty$. Finally, we compute (at least approximately) the global maximum of the main term in the asymptotic formula of the density $\gamma(A_5, S(n_1, n_2, n_3, n_4))$ in the region defined by $0 < n_1, n_2, n_3, n_4 < n$ and $n_1 + n_2 + n_3 + n_4 = n$.

As a closing comment, when we consider five or more even binary trees instead of four in the tree configuration of Figure 6.5, we do not seem to get a better lower bound. We therefore expect our construction to be best possible.

Among the topological trees with fewer than five leaves, the 4-leaf ternary tree Q_4 (Figure 6.1) is the only one for which we are yet to determine an exact inducibility. What is the inducibility of Q_4 (at least in ternary trees)? In what follows, we shall derive a lower and upper bound on $I_3(Q_4)$. Our second main theorem reads as follows:

Theorem 6.2.0.2. For the ternary tree Q_4, we have

$$0.141827 \approx \frac{59}{416} \leq I_3(Q_4) \leq \frac{73848853}{514606225} \approx 0.143506\,.$$

The proof of the lower bound in Theorem 6.2.0.2 is accomplished by an explicit construction (as in Theorem 6.2.0.1), while the upper bound is obtained by means of a computer search. We defer them to Section 6.5.

The star with k leaves is obtained by joining k distinct vertices to a new vertex (the root of the star). We shall denote it with the symbol C_k. The complete d-ary tree of height h is the strictly d-ary tree in which the distance from every leaf to the root is h. Such a tree has d^h leaves in total and shall be denoted with the symbol CD_h^d.

For a positive integer $k \geq 3$, denote by Q_k the tree whose branches are C_{k-1} and C_1 (the single leaf). The following proposition will serve as an intermediary result to proving a new lower bound on the inducibility of the tree Q_4. Its proof will be given in Section 6.5.

Proposition 6.2.0.3. For every positive integer $k \geq 3$, the formula

$$c(Q_k, CD_h^d) = \frac{(d-1)\binom{d}{k-1}}{d^{k-1}-d} \cdot d^h \left(\frac{d^{(k-1)h} - d^{k-1}}{d^{k-1}-1} - \frac{d^h - d}{d-1} \right)$$

holds for every $d \geq 2$ and all $h \geq 1$. In particular, we have

$$I_d(Q_k) \geq \frac{k!(d-1)\binom{d}{k-1}}{(d^{k-1}-d)(d^{k-1}-1)}$$

for every d and $k \geq 3$.

The next proposition shows that the bounds mentioned in Theorems 6.2.0.1 and 6.2.0.2 are much better than the natural bounds provided by the complete d-ary trees, cf. Chapter 5.

Proposition 6.2.0.4. For the trees Q_4 and A_5, we have

$$\lim_{h\to\infty} \gamma\big(Q_4, CD_h^3\big) = \frac{1}{13} \approx 0.076923077$$

and

$$\lim_{h\to\infty} \gamma\big(A_5, CD_h^2\big) = \frac{1}{7} \approx 0.142857143\,.$$

Proof. The specialisation $d = 3$ and $k = 4$ in Proposition 6.2.0.3 yields

$$\lim_{h\to\infty} \gamma\big(Q_4, CD_h^3\big) = \frac{1}{13}\,.$$

As a special case of a result in Theorem 5.3.0.1 (Chapter 5), we know that

$$\lim_{h\to\infty} \gamma\big(A_5, CD_h^2\big) = \frac{2\cdot 5}{2^5 - 2} \cdot I_2\big(CD_2^2\big)\,,$$

while it was proved in the same source (see also [1, Proposition 2]) that $I_2\big(CD_2^2\big) = 3/7$. This completes the proof of the proposition. $\square$

6.3 An algorithm for the maximum

Our next theorem will be used to prove the upper bound on the inducibility of each of the trees A_5 and Q_4. Here, we shall only discuss the tree A_5 (the case of Q_4 is analogous, as will become clear from the proof). We know from Theorem 3.3.0.1 in Chapter 3 that

$$I_d(S) \leq \max_{\substack{|T|=n\\ T\ d\text{-ary tree}}} \gamma(S,T)$$

for all d-ary trees S and $n \geq |S|$. Thus it suffices to determine the value on the right (which can be shown to be decreasing in n as $n \geq |S|$) for as large a value of n as possible to obtain an upper bound. This will be the main goal of this section, where an algorithm for this purpose will be presented. We first need a series of lemmas.

If v is a vertex of a topological tree T, then the subtree $T[v]$ consisting of v and all its descendants in T is called a *fringe* subtree of T. In other words, $T[v]$ is the subtree of T rooted at v.

Lemma 6.3.0.1. Let v be a vertex of a binary tree T, and let $T[v]$ be the fringe subtree rooted at v. The number of copies of A_5 in T can be expressed as

$$c(A_5, T) = c(A_5, T[v]) + (|T| - |T[v]|)c(CD_2^2, T[v]) + R,$$

where R only depends on the size of $T[v]$ (and the rest of T), but not its precise structure.

Proof. If a set of leaves contains at most three leaves of $T[v]$, then there is only one possibility for the tree induced by them inside of $T[v]$. Thus the number of copies of A_5 in T that contain at most three leaves of $T[v]$ only depends on the size of $T[v]$, but not its shape. This leaves us with

- copies of A_5 that are entirely contained in $T[v]$; their number is clearly $c(A_5, T[v])$,
- copies of A_5 that contain precisely four leaves of $T[v]$; there are $|T| - |T[v]|$ other leaves, and the four leaves in $T[v]$ have to induce a copy of CD_2^2 to obtain a copy of A_5. Thus the number of these copies is $(|T| - |T[v]|)c(CD_2^2, T[v])$.

The statement of the lemma follows. $\square$

Lemma 6.3.0.2. Let v be a vertex of a binary tree T, and let $T[v]$ be the fringe subtree rooted at v. Let S be a binary tree of the same size as $T[v]$ that satisfies

$$c(CD_2^2, S) \geq c(CD_2^2, T[v]) \quad \text{and} \quad c(A_5, S) \geq c(A_5, T[v]),$$

at least one of them with strict inequality. Let T' be obtained from T by replacing $T[v]$ with S; then we have

$$c(A_5, T') > c(A_5, T).$$

Proof. This is immediate from the previous lemma. □

Lemma 6.3.0.3. Let v be a vertex of a binary tree T, and let $T[v]$ be the fringe subtree rooted at v. Let S_1 and S_2 be two binary trees of the same size as $T[v]$ that satisfy

$$c(CD_2^2, S_1) > c(CD_2^2, T[v]) > c(CD_2^2, S_2)$$

and

$$c(A_5, S_1) < c(A_5, T[v]) < c(A_5, S_2).$$

Suppose further that

$$\frac{c(A_5, S_1) - c(A_5, T[v])}{c(CD_2^2, S_1) - c(CD_2^2, T[v])} \geq \frac{c(A_5, T[v]) - c(A_5, S_2)}{c(CD_2^2, T[v]) - c(CD_2^2, S_2)}. \tag{6.3.1}$$

Let T_1 and T_2 be obtained from T by replacing $T[v]$ with S_1 and S_2 respectively; then we have

$$\max\left(c(A_5, T_1), c(A_5, T_2)\right) \geq c(A_5, T). \tag{6.3.2}$$

If strict inequality holds in (6.3.1), then we also have strict inequality in (6.3.2).

Proof. Let $k = |T| - |T[v]|$. By Lemma 6.3.0.1, we have

$$\begin{aligned} c(A_5, T_1) - c(A_5, T) &= c(A_5, S_1) - c(A_5, T[v]) + k\big(c(CD_2^2, S_1) - c(CD_2^2, T[v])\big) \\ &= \big(c(CD_2^2, S_1) - c(CD_2^2, T[v])\big) \\ &\qquad \cdot \left(k + \frac{c(A_5, S_1) - c(A_5, T[v])}{c(CD_2^2, S_1) - c(CD_2^2, T[v])}\right). \end{aligned}$$

If

$$\frac{c(A_5, S_1) - c(A_5, T[v])}{c(CD_2^2, S_1) - c(CD_2^2, T[v])} \geq -k,$$

then we are done, since $c(A_5, T_1) \geq c(A_5, T)$. Otherwise, (6.3.1) implies that

$$\frac{c(A_5, T[v]) - c(A_5, S_2)}{c(CD_2^2, T[v]) - c(CD_2^2, S_2)} < -k.$$

Now it follows that

$$\begin{aligned} c(A_5, T_2) - c(A_5, T) &= c(A_5, S_2) - c(A_5, T[v]) + k\big(c(CD_2^2, S_2) - c(CD_2^2, T[v])\big) \\ &= \big(c(CD_2^2, S_2) - c(CD_2^2, T[v])\big) \\ &\qquad \left(k + \frac{c(A_5, T[v]) - c(A_5, S_2)}{c(CD_2^2, T[v]) - c(CD_2^2, S_2)}\right) > 0, \end{aligned}$$

so $c(A_5, T_2) \geq c(A_5, T)$. Either way, we have (6.3.2). Equality can only hold if both quotients in (6.3.1) are equal to $-k$. This completes the proof. □

Lemma 6.3.0.4. Let v be a vertex of a binary tree T, and let $T[v]$ be the fringe subtree rooted at v. Let S be a binary tree of the same size as $T[v]$ that satisfies

$$c(CD_2^2, S) > c(CD_2^2, T[v])$$

and

$$c(A_5, S) < c(A_5, T[v]).$$

Suppose further that

$$\frac{c(A_5, S) - c(A_5, T[v])}{c(CD_2^2, S) - c(CD_2^2, T[v])} \geq |T[v]| - |T|. \tag{6.3.3}$$

Let T' be obtained from T by replacing $T[v]$ with S; then we have

$$c(A_5, T') \geq c(A_5, T). \tag{6.3.4}$$

If strict inequality holds in (6.3.3), then we also have strict inequality in (6.3.4).

Proof. As in the proof of the previous lemma, we have

$$c(A_5, T') - c(A_5, T) = \left(c(CD_2^2, S) - c(CD_2^2, T[v])\right) \left(|T| - |T[v]| + \frac{c(A_5, S) - c(A_5, T[v])}{c(CD_2^2, S) - c(CD_2^2, T[v])}\right).$$

The statement follows immediately. □

Now we are ready to describe the algorithm to determine the maximum number of copies of A_5 in a binary tree with n leaves. To this end, we define a sequence of sets of binary trees: intuitively speaking, $\mathcal{L}(n)$ consists of trees with n leaves that can potentially occur as fringe subtrees of "optimal" trees, i.e., binary trees that maximize the number of copies of A_5. A formal recursive definition will be provided below. We also associate every tree T with the pair $P(T) = (c(A_5, T), c(CD_2^2, T))$, which can be interpreted as a point in the plane, and we set

$$L(n) = \{P(T) : T \in \mathcal{L}(n)\}$$

for every n. The sets $\mathcal{L}(n)$ are recursively defined as follows:

1. The set $\mathcal{L}(1)$ only consists of one tree, which only has a single vertex.

2. For $n > 1$, we consider all binary trees with n leaves for which each branch lies in one of the sets $\mathcal{L}(m)$ for some $m < n$. Clearly, if one branch lies in $\mathcal{L}(k)$, the other has to lie in $\mathcal{L}(n-k)$. For reasons to become clear later (essentially, we are applying Lemma 6.3.0.4), we will be even more restrictive: we consider all binary trees with n leaves whose branches both lie in

$$\bigcup_{m<n} \Big\{ T \in \mathcal{L}(m) \,:\, \text{there is no } S \in \mathcal{L}(m) \text{ such that} \\ c(CD_2^2, S) > c(CD_2^2, T), c(A_5, S) < c(A_5, T), \text{ and} \\ \frac{c(A_5, S) - c(A_5, T[v])}{c(CD_2^2, S) - c(CD_2^2, T)} \geq m - n \Big\}.$$

   This gives us a preliminary set $\mathcal{H}_1(n)$.

3. If there are two trees T and T' in $\mathcal{H}_1(n)$ such that

$$c(CD_2^2, T) \geq c(CD_2^2, T') \qquad \text{and} \qquad c(A_5, T) \geq c(A_5, T'),$$

   remove T' from $\mathcal{H}_1(n)$. If we have equality in both inequalities, we can arbitrarily remove either T or T'. In geometric terms, the condition means that the point $P(T')$ lies to the left and below the point $P(T)$ in the plane. We repeat this step until there are no two trees T and T' satisfying the aforementioned condition anymore. At the end, we are left with a set $\mathcal{H}_2(n)$.

4. As a final reduction step, we eliminate all trees T from $\mathcal{H}_2(n)$ for which there exist two trees S_1 and S_2 in $\mathcal{H}_2(n)$ such that the inequalities of Lemma 6.3.0.3 hold, i.e.,

$$c(CD_2^2, S_1) > c(CD_2^2, T) > c(CD_2^2, S_2)$$

   and

$$c(A_5, S_1) < c(A_5, T) < c(A_5, S_2)$$

   as well as

$$\frac{c(A_5, S_1) - c(A_5, T)}{c(CD_2^2, S_1) - c(CD_2^2, T)} \geq \frac{c(A_5, T) - c(A_5, S_2)}{c(CD_2^2, T) - c(CD_2^2, S_2)}.$$

Considering the set of points $\{P(T) : T \in \mathcal{H}_2(T)\}$ in the plane, this amounts to taking the upper envelope of the points. The resulting set after this reduction is $\mathcal{L}(n)$. At this point, we can arrange the elements of $\mathcal{L}(n)$ as a list of trees $T_1, T_2, \ldots, T_r$ such that

$$c(CD_2^2, T_1) < c(CD_2^2, T_2) < \cdots < c(CD_2^2, T_r),$$

$$c(A_5, T_1) > c(A_5, T_2) > \cdots > c(A_5, T_r),$$

and the sequence of "slopes"

$$\frac{c(A_5, T_{j+1}) - c(A_5, T_j)}{c(CD_2^2, T_{j+1}) - c(CD_2^2, T_j)}$$

is strictly decreasing. This also makes it easier to construct the set in step (2): the trees from $\mathcal{L}(m)$ that are allowed as branches are precisely those starting from the point where the slope is less than $m - n$.

Due to the rules of the two elimination steps, the following holds for all $T \in \mathcal{H}_1(n)$ at the end:

- Either there exists an $S \in \mathcal{L}(n)$ (possibly $T = S$) such that

$$c(CD_2^2, S) \geq c(CD_2^2, T) \qquad \text{and} \qquad c(A_5, S) \geq c(A_5, T),$$

- or there exist two trees $S_1, S_2 \in \mathcal{L}(n)$ such that

$$\begin{aligned} &c(CD_2^2, S_1) > c(CD_2^2, T) > c(CD_2^2, S_2), \\ &c(A_5, S_1) < c(A_5, T) < c(A_5, S_2) \end{aligned}$$

  and

$$\frac{c(A_5, S_1) - c(A_5, T)}{c(CD_2^2, S_1) - c(CD_2^2, T)} \geq \frac{c(A_5, T) - c(A_5, S_2)}{c(CD_2^2, T) - c(CD_2^2, S_2)}.$$

The following theorem shows that the maximum of $c(A_5, T)$ for binary trees T with a given number of leaves can be determined purely by focusing on the sets $\mathcal{L}(n)$.

Theorem 6.3.0.5. For every positive integer n, there exists a binary tree M_n with n leaves such that

$$c(A_5, M_n) = \max_{\substack{|T|=n \\ T \text{ binary tree}}} c(A_5, T)$$

and all fringe subtrees of M_n (including M_n itself) lie in $\bigcup_{k\geq 1} \mathcal{L}(k)$. In particular,

$$\max_{\substack{|T|=n \\ T \text{ binary tree}}} c(A_5, T) = \max_{T\in\mathcal{L}(n)} c(A_5, T).$$

Proof. Suppose that the statement does not hold, and let m be minimal with the property that there is a positive integer n such that every "optimal" tree (tree attaining the maximum $\max_{|T|=n} c(A_5, T)$) has a fringe subtree with m or fewer leaves that does not lie in $\bigcup_{1\leq k\leq m} \mathcal{L}(k)$. Clearly, $m > 1$.

By our choice of m, there must be an optimal tree T with n leaves for which all fringe subtrees with less than m leaves lie in $\bigcup_{1\leq k<m} \mathcal{L}(k)$. Among all possible choices of T, we can choose one for which the number of m-leaf fringe subtrees that do not lie in $\mathcal{L}(m)$ is minimal. Consider one of these fringe subtrees $T[v]$. Both its branches lie in $\bigcup_{1\leq k<m} \mathcal{L}(k)$, which leaves us with the following possible reasons why $T[v]$ is not in $\mathcal{L}(m)$:

- There is a binary tree $S \in \mathcal{L}(m)$ such that

  $$c(CD_2^2, S) \geq c(CD_2^2, T[v]) \quad \text{and} \quad c(A_5, S) \geq c(A_5, T[v]).$$

  In this case, we can replace $T[v]$ by S to obtain a new tree with at least as many copies of A_5 as T by Lemma 6.3.0.2. This contradicts our choice of T (it is either not optimal, or it does not have the smallest number of m-leaf fringe subtrees that do not lie in $\mathcal{L}(m)$).

- There are binary trees $S_1, S_2 \in \mathcal{L}(m)$ such that

  $$c(CD_2^2, S_1) > c(CD_2^2, T) > c(CD_2^2, S_2),\ c(A_5, S_1) < c(A_5, T) < c(A_5, S_2)$$

  and
  $$\frac{c(A_5, S_1) - c(A_5, T)}{c(CD_2^2, S_1) - c(CD_2^2, T)} \geq \frac{c(A_5, T) - c(A_5, S_2)}{c(CD_2^2, T) - c(CD_2^2, S_2)}.$$

In this case, we can replace $T[v]$ by either S_1 or S_2 to obtain a contradiction in the same way as in the previous case (now by means of Lemma 6.3.0.3).

- The branches of $T[v]$ do not satisfy the condition of step (2) in the construction of $\mathcal{L}(n)$. Suppose that for one of the branches B, there is a tree S in $\mathcal{L}_{|B|}$ such that

$$c(CD_2^2, S) > c(CD_2^2, B), c(A_5, S) < c(A_5, B),$$

and

$$\frac{c(A_5, S) - c(A_5, B)}{c(CD_2^2, S) - c(CD_2^2, B)} \geq |B| - |T[v]| \geq |B| - |T|.$$

We can replace B by S, and do likewise with the other branch of S if necessary. In this way, $T[v]$ is replaced by a tree in $\mathcal{L}(m)$, and Lemma 6.3.0.4 now yields the desired contradiction.

Since we reach a contradiction in all possible cases, the proof is complete. □

For a practical implementation of this algorithm, it actually suffices to work with the lists

$$L(n) = \{P(T) : T \in \mathcal{L}(n)\}$$

that contain the values of $P(T) = (c(A_5, T), c(CD_2^2, T))$. These values can be calculated recursively: if the branches of a binary tree T are B_1 and B_2, we have

$$c(A_5, T) = c(A_5, B_1) + c(A_5, B_2) + |B_1| c(CD_2^2, B_2) + |B_2| c(CD_2^2, B_1) \quad (6.3.5)$$

and

$$c(CD_2^2, T) = c(CD_2^2, B_1) + c(CD_2^2, B_2) + \binom{|B_1|}{2}\binom{|B_2|}{2}. \quad (6.3.6)$$

They can be explained as follows:

- A subset of five leaves of the leaf-set of T can either be a subset of leaves of B_1, or a subset of leaves of B_2, or splits into leaves of both B_1 and B_2. In the latter case, the split must be of the type $1-4$ (or

$4-1$) as the branches of A_5 are C_1 and CD_2^2. Moreover, the root of the leaf-induced subtree in this case coincides with the root of the tree T. This proves the recursion for A_5.

- The four leaves of T that induce the tree CD_2^2 can either lie entirely in T_1 or T_2; or precisely two leaves of each of the branches B_1 and B_2 of T induce the star C_2 to obtain a copy of CD_2^2. This proves the recursive formula for CD_2^2.

Thus it is never necessary to store full tree structures. At the end, the maximum

$$\max_{\substack{|T|=n \\ T \text{ binary tree}}} c(A_5, T)$$

can be determined easily from $L(n)$.

6.4 Proof of Theorem 6.2.0.1

This section is devoted to proving Theorem 6.2.0.1. Recall that we are going to use the binary tree $S(n_1, n_2, n_3, n_4)$ presented in Figure 6.5. Moreover, we now need to consider only $I_2(A_5)$ because it is established in Corollary 4.2.0.8 (Chapter 4) that $J(B) = I_2(B)$ for every binary tree B.

Proof of Theorem 6.2.0.1. Let us set $n = n_1 + n_2 + n_3 + n_4$. Recall from equation (6.3.5) that a recursion for the number of copies of A_5 in any binary tree T with branches B_1 and B_2 is given by

$$c(A_5, T) = c(A_5, B_1) + c(A_5, B_2) + |B_1| \cdot c\big(CD_2^2, B_2\big) + |B_2| \cdot c\big(CD_2^2, B_1\big)\,.$$

So for the tree $S(n_1, n_2, n_3, n_4)$, we obtain

$$\begin{aligned} c\big(A_5, S(n_1, n_2, n_3, n_4)\big) &= c\big(A_5, E^2_{n_1}\big) + c\big(A_5, E^2_{n_2}\big) + c\big(A_5, E^2_{n_3}\big) + c\big(A_5, E^2_{n_4}\big) \\ &\quad + n_3 \cdot c\big(CD_2^2, E^2_{n_4}\big) + n_4 \cdot c\big(CD_2^2, E^2_{n_3}\big) \\ &\quad + n_2 \cdot c\big(CD_2^2, T_{n_3,n_4}\big) + (n_3 + n_4) \cdot c\big(CD_2^2, E^2_{n_2}\big) \\ &\quad + n_1 \cdot c\big(CD_2^2, T_{n_2,n_3,n_4}\big) + (n_2 + n_3 + n_4) \cdot c\big(CD_2^2, E^2_{n_1}\big)\,, \end{aligned} \tag{6.4.1}$$

where T_{n_3,n_4} is the binary tree whose branches are the even binary trees $E^2_{n_3}$ and $E^2_{n_4}$, while T_{n_2,n_3,n_4} is the binary tree whose branches are $E^2_{n_2}$ and T_{n_3,n_4}.

Also, recall from equation (6.3.6) that a recursion for the number of copies of CD^2_2 in any binary tree T with branches B_1 and B_2 is given by

$$c(CD^2_2, T) = c(CD^2_2, B_1) + c(CD^2_2, B_2) + \binom{|B_1|}{2}\binom{|B_2|}{2}.$$

So for the binary tree T_{n_2,n_3,n_4}, we get

$$c\big(CD^2_2, T_{n_2,n_3,n_4}\big) = c\big(CD^2_2, E^2_{n_2}\big) + c\big(CD^2_2, T_{n_3,n_4}\big) + \binom{n_2}{2}\binom{n_3+n_4}{2}.$$

Likewise,

$$c\big(CD^2_2, T_{n_3,n_4}\big) = c\big(CD^2_2, E^2_{n_3}\big) + c\big(CD^2_2, E^2_{n_4}\big) + \binom{n_3}{2}\binom{n_4}{2}.$$

Thus, equation (6.4.1) becomes

$$\begin{aligned} c\big(A_5, S(n_1,n_2,n_3,n_4)\big) &= c\big(A_5, E^2_{n_1}\big) + c\big(A_5, E^2_{n_2}\big) + c\big(A_5, E^2_{n_3}\big) + c\big(A_5, E^2_{n_4}\big) \\ &\quad + n_3 \cdot c\big(CD^2_2, E^2_{n_4}\big) + n_4 \cdot c\big(CD^2_2, E^2_{n_3}\big) + (n_3+n_4)\cdot c\big(CD^2_2, E^2_{n_2}\big) \\ &\quad + n_2\Big(c\big(CD^2_2, E^2_{n_3}\big) + c\big(CD^2_2, E^2_{n_4}\big) + \binom{n_3}{2}\binom{n_4}{2}\Big) \\ &\quad + n_1\Big(c\big(CD^2_2, E^2_{n_2}\big) + c\big(CD^2_2, E^2_{n_3}\big) + c\big(CD^2_2, E^2_{n_4}\big) + \binom{n_3}{2}\binom{n_4}{2} \\ &\quad + \binom{n_2}{2}\binom{n_3+n_4}{2}\Big) + (n_2+n_3+n_4)\cdot c\big(CD^2_2, E^2_{n_1}\big) \end{aligned}$$

after combining everything. As a special case of Theorem 5.4.0.7 in Chapter 5, we have

$$c\big(CD^2_2, E^2_n\big) = \frac{1}{56}\cdot n^4 + \mathcal{O}(n^3)$$

for all n. On the other hand, using this asymptotic formula along with the recursion

$$\begin{aligned} c\big(A_5, E^2_n\big) &= c\big(A_5, E^2_{\lfloor n/2 \rfloor}\big) + c\big(A_5, E^2_{\lceil n/2 \rceil}\big) \\ &\qquad + \lfloor n/2 \rfloor \cdot c\big(CD^2_2, E^2_{\lceil n/2 \rceil}\big) + \lceil n/2 \rceil \cdot c\big(CD^2_2, E^2_{\lfloor n/2 \rfloor}\big), \end{aligned}$$

which follows from (6.3.5) by the definition of the even binary tree E_n^2, it is not hard to prove that there exist absolute constants $K_1, K_2 \geq 0$ such that the double inequality

$$\frac{1}{840} \cdot n^5 - K_1 \cdot n^4 \leq c\big(A_5, E_n^2\big) \leq \frac{1}{840} \cdot n^5 + K_2 \cdot n^4$$

holds for all n—the details are omitted. Moreover, we have the asymptotic formulas

$$c\big(CD_2^2, E_n^2\big) = \frac{1}{56} \cdot n^4 + \mathcal{O}(n^3) \quad \text{and} \quad c\big(A_5, E_n^2\big) = \frac{1}{840} \cdot n^5 + \mathcal{O}(n^4)\,.$$

Now, let x_1, x_2, x_3, x_4 be positive real numbers with $x_1 + x_2 + x_3 + x_4 = 1$. We set $n_i = \lfloor x_i n \rfloor = x_i n + \mathcal{O}(1)$ for $i = 1, 2, 3, 4$. Combining all the asymptotic formulas, we we can now rewrite $c\big(A_5, S(n_1, n_2, n_3, n_4)\big)$ as follows:

$$\begin{aligned}
c\big(A_5, S(n_1, n_2, n_3, n_4)\big) &= \frac{n^5}{840}\big(x_1^5 + x_2^5 + x_3^5 + x_4^5\big) + \frac{n^5}{56}\big(x_3 \cdot x_4^4 + x_4 \cdot x_3^4\big) \\
&\quad + \frac{n^5}{56} \cdot x_2\big(x_3^4 + x_4^4 + 14 \cdot x_3^2 \cdot x_4^2\big) + \frac{n^5}{56}(x_3 + x_4)x_2^4 \\
&\quad + \frac{n^5}{56} \cdot x_1\Big(x_2^4 + x_3^4 + x_4^4 + 14 \cdot x_3^2 \cdot x_4^2 + 14 \cdot x_2^2(x_3 + x_4)^2\Big) \\
&\quad + \frac{n^5}{56}(x_2 + x_3 + x_4)x_1^4 + \mathcal{O}(n^4)\,.
\end{aligned}$$

Set

$$\begin{aligned}
F(x_1, x_2, x_3) &= \frac{1}{840}\big(x_1^5 + x_2^5 + x_3^5 + (1 - x_1 - x_2 - x_3)^5\big) \\
&\quad + \frac{1}{56}\Big(x_3(1 - x_1 - x_2 - x_3)^4 + (1 - x_1 - x_2 - x_3)x_3^4 \\
&\quad + x_2\big(x_3^4 + (1 - x_1 - x_2 - x_3)^4 + 14 \cdot x_3^2(1 - x_1 - x_2 - x_3)^2\big) \\
&\quad + (1 - x_1 - x_2)x_2^4 + x_1\Big(x_2^4 + x_3^4 + (1 - x_1 - x_2 - x_3)^4 \\
&\quad + 14 \cdot x_3^2(1 - x_1 - x_2 - x_3)^2 + 14 \cdot x_2^2(1 - x_1 - x_2)^2\Big) + \frac{1}{56}(1 - x_1)x_1^4\Big)\,.
\end{aligned}$$

Then we obtain

$$c\big(A_5, S(n_1, n_2, n_3, n_4)\big) = F(x_1, x_2, x_3) \cdot n^5 + \mathcal{O}(n^4)$$

as $x_1 + x_2 + x_3 + x_4 = 1$. With the help of a computer, we find that the global maximum of the function $F(x_1, x_2, x_3)$ in the region covered by the inequalities $0 < x_1, x_2, x_3 < 1,\ x_1 + x_2 + x_3 < 1$ is approximately attained at the points

$$(x_1 = 0.0253477306003, x_2 = 0.0514257548195463, x_3 = 0.788023107375217)$$

and

$$(x_1 = 0.0253477306003, x_2 = 0.0514257548195463, x_3 = 0.135203390298812).$$

Thus we have

$$\begin{aligned} &\max_{\substack{0<x_1,x_2,x_3<1\\ x_1+x_2+x_3<1}} F(x_1,x_2,x_3) \geq \\ &\qquad F(0.0253477306003, 0.0514257548195463, 0.788023107375217) \\ &\qquad = 0.0020589291815410\,. \end{aligned}$$

The inequality

$$I_2(A_5) \geq 0.00205892918154170 \times 5! \approx 0.247071501785004$$

follows. This concludes the proof of the first part of the theorem.

For the upper bound, we make use of Theorem 6.3.0.5 which states that the maximum of $c(A_5, T)$ for binary trees T with n leaves can be determined purely by focusing on the sets $\mathcal{L}(n)$ whose algorithmic description is given in Section 6.3. Recall that by Theorem 3.3.0.1 in Chapter 3, we have

$$I_2(A_5) \leq \max_{\substack{|T|=n\\ T \text{ binary tree}}} \gamma(A_5, T)$$

for every $n \geq 5$. Thus we want to calculate the maximum for different values of n. When our algorithm terminates, the maximum number of copies of A_5 among all binary trees with n leaves can be read off as the greatest x-coordinate (first coordinate) of the elements of $L(n)$, that is the x-coordinate of the very first element of $L(n)$—see the discussion before Theorem 6.3.0.5.

We have implemented this algorithm in Mathematica. The notebook can be accessed at `http://math.sun.ac.za/~swagner/TreeA5Final`. The precise values of

$$a_n = \max_{\substack{|T|=n\\ T \text{ binary tree}}} \gamma(A_5, T)$$

Table 6.1: Maximum density a_n of A_5 among n-leaf binary trees

n	5	6	7	8	9	10	20	30
a_n	1	$\frac{1}{2}$	$\frac{3}{7}$	$\frac{11}{28}$	$\frac{23}{63}$	$\frac{1}{3}$	$\frac{553}{1938}$	$\frac{19219}{71253}$
n	40	50	60	70	80	90	100	150
a_n	$\frac{57793}{219336}$	$\frac{550621}{2118760}$	$\frac{351943}{1365378}$	$\frac{44899}{175406}$	$\frac{6127045}{24040016}$	$\frac{930032}{3662439}$	$\frac{3177631}{12547920}$	$\frac{24765738}{98600005}$
n	200	250	300	350	400	500	600	700
a_n	0.250153	0.249543	0.249142	0.248854	0.24864	0.24834	0.248143	0.248001
n	800	900	1000	1200	1400	1600	1800	2000
a_n	0.247894	0.247812	0.247747	0.247648	0.247577	0.247524	0.247483	0.24745

have been computed for $n \leq 2000$—see Table 6.1 .

It follows that $I_2(A_5) \leq a_{2000} \approx 0.24745$. This completes the proof of the theorem. □

6.5 Proof of Theorem 6.2.0.2

Let us first provide a proof of Proposition 6.2.0.3, which is an intermediary step in the proof of Theorem 6.2.0.2:

Proof of Proposition 6.2.0.3. Let $k \geq 3$ and $d \geq k-1$ be fixed; note that

$$c\big(Q_k, CD_h^d\big) = 0 \ \text{ for } \ d < k-1\,.$$

For $h = 1$, we have $c\big(Q_k, CD_h^d\big) = c(Q_k, C_d) = 0$. Since for the case $h = 1$, the statement holds trivially, we can safely assume $h \geq 2$ and proceed by induction on h.

We distinguish possible cases that can happen for a subset of k leaves of the tree CD_h^d:

- all k leaves belong to the same branch of CD_h^d. The total number of these subsets of leaves that induce the tree Q_k is given by $d \cdot c\big(Q_k, CD_{h-1}^d\big)$, as all the branches of CD_h^d are isomorphic to CD_{h-1}^d,
- more than two of the branches of CD_h^d contain at least one of the k leaves. In this case the leaf-induced subtree is not isomorphic to Q_k (as the root degree of Q_k is 2),

- exactly two of the branches of CD_h^d contain at least two of the k leaves each. In this case the leaf-induced subtree is not isomorphic to Q_k (as one of the branches of Q_k is the single leaf),
- one branch of CD_h^d contains exactly one of the leaves and another branch of CD_h^d contains $k-1$ leaves. Since $k > 2$, the total number of these subsets of leaves that induce the tree Q_k is given by $2 \cdot d^{h-1} \cdot c(C_{k-1}, CD_{h-1}^d)$ for every choice of two branches of CD_h^d.

Therefore, a recursion for $c(Q_k, CD_h^d)$ is given by

$$\begin{aligned} c(Q_k, CD_h^d) &= d \cdot c(Q_k, CD_{h-1}^d) + 2\binom{d}{2} \cdot d^{h-1} \cdot c(C_{k-1}, CD_{h-1}^d) \\ &= d \cdot c(Q_k, CD_{h-1}^d) + (d-1)d^h \binom{d}{k-1} \left(\frac{d^{(k-1)(h-1)} - d^{h-1}}{d^{k-1} - d} \right), \end{aligned}$$

where the last step uses the identity

$$c(C_k, CD_h^d) = \binom{d}{k} \frac{d^{k \cdot h} - d^h}{d^k - d} \tag{6.5.1}$$

valid for every $k \geq 2$ – formula (6.5.1) can be found in the proof of Theorem 3.2.0.1 (Chapter 3). The induction hypothesis (with respect to h) gives

$$\begin{aligned} c(Q_k, CD_h^d) = & \frac{(d-1)\binom{d}{k-1}}{d^{k-1} - d} \cdot d^h \left(\frac{d^{(k-1)(h-1)} - d^{k-1}}{d^{k-1} - 1} - \frac{d^{h-1} - d}{d - 1} \right) \\ & + (d-1)d^h \binom{d}{k-1} \left(\frac{d^{(k-1)\cdot(h-1)} - d^{h-1}}{d^{k-1} - d} \right) \\ = & \frac{(d-1)\binom{d}{k-1}}{d^{k-1} - d} \cdot d^h \left(d^{(k-1)(h-1)} + \frac{d^{(k-1)(h-1)} - d^{k-1}}{d^{k-1} - 1} \right. \\ & \left. - \frac{d^{h-1} - d}{d - 1} - d^{h-1} \right), \end{aligned}$$

which, after simplification, yields the desired equality. The statement on the inducibility follows by passing to the limit of the density $\gamma(Q_k, CD_h^d)$ as $h \to \infty$:

$$I_d(Q_k) \geq \lim_{h \to \infty} \gamma(Q_k, CD_h^d) = \frac{k!(d-1)\binom{d}{k-1}}{(d^{k-1} - d)(d^{k-1} - 1)}.$$

□

We can now focus on Theorem 6.2.0.2.

Proof of Theorem 6.2.0.2. First off, we construct a new family of ternary trees: given a nonnegative integer $h \geq 0$, attach one copy of each of the complete ternary trees CD_h^3 and CD_{h+1}^3 to a common vertex (their respective roots are joined to a new vertex) to form a ternary tree which we shall name W_h^3. For example, W_0^3 is the tree Q_4. See also Figure 6.6 for the ternary tree W_1^3.

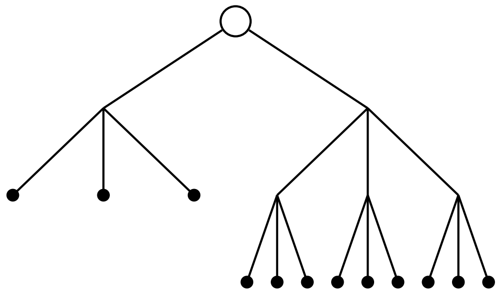


Figure 6.6: The ternary tree W_1^3 defined in the proof of Theorem 6.2.0.2.

Let us prove that

$$\lim_{h\to\infty} \gamma(Q_4, W_h^3) = 59/416\,.$$

To justify the specific choice, let us consider a more general construction. For positive integers n_1 and n_2, we consider the ternary tree (which we simply denote by T_{n_1,n_2}) whose branches are even ternary trees with n_1 and n_2 leaves, respectively. The even ternary tree E_n^3 with n leaves is obtained recursively as follows: E_1^3 is the tree with only one vertex; E_2^3 is the star with two leaves; for $n > 2$, the branches of E_n^3 are the even ternary trees $E_{k_1}^3$, $E_{k_2}^2$ and $E_{k_3}^3$ with k_1, k_2, k_3 as equal as possible and $k_1 + k_2 + k_3 = n$.

According to Proposition 6.2.0.3, we have

$$c(Q_4, CD_h^d) = \frac{(d-2)d^{4\cdot h}}{6(d+1)(d^2+d+1)} + \mathcal{O}(d^{2\cdot h})$$

for every d and all $h \geq 3$. In particular, the asymptotic formula

$$c(Q_4, CD_h^3) = \frac{1}{312} \cdot 3^{4\cdot h} + \mathcal{O}(3^{2\cdot h})$$

is obtained for all $h \geq 3$. Using the recursion

$$\begin{aligned} c\big(Q_4, E_n^3\big) &= c\big(Q_4, E_{k_1}^3\big) + c\big(Q_4, E_{k_2}^3\big) + c\big(Q_4, E_{k_3}^3\big) \\ &\quad + k_1 \cdot c\big(C_3, E_{k_2}^3\big) + k_2 \cdot c\big(C_3, E_{k_1}^3\big) + k_1 \cdot c\big(C_3, E_{k_3}^3\big) \\ &\quad + k_3 \cdot c\big(C_3, E_{k_1}^3\big) + k_2 \cdot c\big(C_3, E_{k_3}^3\big) + k_3 \cdot c\big(C_3, E_{k_2}^3\big)\,, \end{aligned}$$

it is not difficult to show that we also have

$$c\big(Q_4, E_n^3\big) = \frac{1}{312} \cdot n^4 + \mathcal{O}(n^3)$$

for all n. On the other hand, we recall that the specialisation $k = 3$ in equation (6.5.1) of the proof of Proposition 6.2.0.3 gives

$$c(C_3, CD_h^3) = \frac{1}{24} \cdot 3^{3h} + \mathcal{O}(3^h)$$

for all $h \geq 1$, and employing

$$c(C_3, E_n^3) = c(C_3, E_{k_1}^3) + c(C_3, E_{k_2}^3) + c(C_3, E_{k_3}^3) + k_1 \cdot k_2 \cdot k_3\,,$$

it is not difficult to show that

$$c(C_3, E_n^3) = \frac{1}{24} \cdot n^3 + \mathcal{O}(n^2)$$

for all n. Moreover, the number of copies of Q_4 in any topological tree T with two branches T_1, T_2 is given by

$$c(Q_4, T) = c(Q_4, T_1) + c(Q_4, T_2) + |T_1| \cdot c(C_3, T_2) + |T_2| \cdot c(C_3, T_1)\,.$$

For $x \in (0,1)$, set $n_1 = \lfloor xn \rfloor$ and $n_2 = \lfloor (1-x)n \rfloor$, and let $n \to \infty$. Combining all the formulas above, we see that an asymptotic formula for $c\big(Q_4, T_{n_1,n_2}\big)$ is given by

$$\begin{aligned} c\big(Q_4, T_{n_1,n_2}\big) &= \frac{1}{312}(x \cdot n)^4 + \frac{1}{312}\big((1-x)n\big)^4 \\ &\quad + x \cdot n \cdot \frac{1}{24}\big((1-x)n\big)^3 + (1-x)n \cdot \frac{1}{24}(x \cdot n)^3 + \mathcal{O}(n^3) \\ &= \frac{1}{312}\big(x^4 + (1-x)^4\big)n^4 + \frac{1}{24}\big(x(1-x)^3 + (1-x)x^3\big)n^4 \\ &\quad + \mathcal{O}(n^3) \\ &= \frac{1}{312}\big(1 + 9x - 33x^2 + 48x^3 - 24x^4\big)n^4 + \mathcal{O}(n^3)\,. \end{aligned}$$

Set

$$f(x) = \frac{1}{312}\left(1 + 9x - 33x^2 + 48x^3 - 24x^4\right).$$

The first derivative of this function is given by

$$f'(x) = \frac{-(2x-1)(4x-3)(4x-1)}{104}.$$

We see that $f(x)$ attains its maximum at $x = 1/4$ (or $x = 3/4$):

$$f(x) \le f\left(\frac{1}{4}\right) = \frac{59}{9984}$$

for all $x \in (0, 1)$. This motivates the choice of the trees W_h^3 defined before. We have

$$I_3(Q_4) \ge \lim_{h\to\infty} \gamma(Q_4, W_h^3) = \frac{4! \cdot 59}{9984} = \frac{59}{416}.$$

This completes the proof of the lower bound in the theorem.

The proof of the upper bound is also via an algorithmic approach and is quite similar to the one given for the binary tree A_5 in Section 6.3. Recall again that by Theorem 3.3.0.1 in Chapter 3,

$$I_3(Q_4) \le \max_{\substack{|T|=n \\ T \text{ ternary tree}}} \gamma(Q_4, T),$$

so the aim is to compute the right hand side for different values of n. The algorithm is essentially the same as for A_5, with the trees Q_4 and C_3 assuming the roles of A_5 and CD_2^2 respectively. The only difference is that trees with two or three branches have to be considered in the construction of the sets $\mathcal{L}(n)$.

For the recursive calculation of $c(Q_4, T)$ and $c(C_3, T)$, we have the formulas

$$\begin{aligned} c(Q_4, T) &= c(Q_4, T_1) + c(Q_4, T_2) + c(Q_4, T_3) + |T_1| \cdot c(C_3, T_2) + |T_2| \cdot c(C_3, T_1) \\ &\quad + |T_1| \cdot c(C_3, T_3) + |T_3| \cdot c(C_3, T_1) + |T_2| \cdot c(C_3, T_3) + |T_3| \cdot c(C_3, T_2) \end{aligned}$$

and

$$c(C_3, T) = c(C_3, T_1) + c(C_3, T_2) + c(C_3, T_3) + |T_1| \cdot |T_2| \cdot |T_3|,$$

where T_1, T_2, T_3 are the branches of T. If there are only two branches, all terms involving T_3 can simply be left out.

Again, we have implemented the algorithm in Mathematica—the notebook can be found at `http://math.sun.ac.za/~swagner/TreeQ4Final`. The exact values of

$$b_n = \max_{\substack{|T|=n \\ T \text{ ternary tree}}} \gamma(Q_4, T)$$

have been determined for values of n up to 500; see Table 6.2.

Table 6.2: Maximum density b_n of Q_4 among n-leaf ternary trees

n	4	5	6	7	8	9	10
$\boldsymbol{b_n}$	1	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{2}{7}$	$\frac{19}{70}$	$\frac{5}{21}$	$\frac{5}{21}$
n	15	20	25	30	35	40	45
$\boldsymbol{b_n}$	$\frac{18}{91}$	$\frac{291}{1615}$	$\frac{1103}{6325}$	$\frac{172}{1015}$	$\frac{1097}{6545}$	$\frac{7452}{45695}$	$\frac{7948}{49665}$
n	50	60	70	80	90	100	150
$\boldsymbol{b_n}$	0.158072	0.155422	0.153588	0.152096	0.150978	0.150264	0.147342
n	200	250	300	350	400	450	500
$\boldsymbol{b_n}$	0.145967	0.145195	0.144651	0.144239	0.143931	0.143691	0.143506

We conclude that

$$I_3(Q_4) \leq b_{500} = \frac{73848853}{514606225} \approx 0.143506\,.$$

This completes the proof of the theorem. □

Chapter 7

The minimum asymptotic density of binary caterpillars

Given a positive integer $d \geq 2$ and two (rooted) d-ary trees D and T such that D has k leaves, the density $\gamma(D,T)$ of D in T is the proportion of all k-element subsets of leaves of T that induce a tree isomorphic to D, after erasing all vertices of outdegree 1. In Chapter 3, it was proved (in an implicit manner) that the limit inferior of this density as the size of T grows to infinity is always zero unless D is the k-leaf binary caterpillar F_k^2 (the binary tree with the property that a path remains upon removal of all the k leaves). Our main theorem in this chapter is an exact formula (involving both d and k) for the limit inferior of $\gamma(F_k^2, T)$ as the size of T tends to infinity.

The material is based on the following paper [36]: *The minimum asymptotic density of binary caterpillars. A. A. V. Dossou-Olory. Preprint* `https://arxiv.org/abs/1804.05731`

7.1 Preliminaries and statement of the main results

In a recent paper [34] (whose results are also presented in Chapter 3), Czabarka, Székely, Wagner and the author of this thesis investigated the inducibility of d-ary trees, which can be thought of as the *maximum asymptotic* *density* of a d-ary tree occurring as a subtree induced by leaves of another d-ary tree with sufficiently large number of leaves.

The inducibility of a d-ary tree (see Chapter 3) is defined as being the maximum asymptotic density of D. Formally speaking, the inducibility $I_d(D)$ of a d-ary tree D is the limit superior of the density of D in T as the number of leaves of T grows to infinity. One of the principal results in Chapter 3 is that

$$I_d(D) = \limsup_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(D,T) = \lim_{n\to\infty} \max_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T)\,. \tag{7.1.1}$$

For the purposes of this chapter, let us define and call the quantity

$$\liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(D,T)$$

the *minimum asymptotic density* of the d-ary tree D in d-ary trees. To put it another way, we mean

$$\liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(D,T) = \lim_{n\to\infty} \min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(D,T)\,,$$

where the proof of existence of the limit is analogous to that in (7.1.1).

An important and recurring theme that appears throughout extremal graph theory is finding the minimum or maximum value of a given graph invariant within a class of graphs all sharing a certain property. Understanding an invariant provides information about the structure of a graph. In particular, the problem of characterising the extremal graphs has been and continues to be a topic of a great interest to graph theorists. It is therefore natural to consider the problem of determining the minimum asymptotic density of a d-ary tree D in d-ary trees. Quite fascinatingly however, it turns out that the study of this problem reduces to the study of the minimum asymptotic density of binary caterpillars.

We recall that a binary caterpillar is a binary tree with the property that its non-leaf vertices form a path starting at the root. We denote by F_k^2 the binary caterpillar with k leaves – see Figure 7.1 for the binary caterpillar with five leaves.

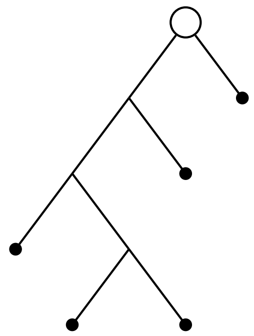


Figure 7.1: The 5-leaf binary caterpillar F_5^2.

The following fundamental result from Chapter 3 characterises all the d-ary trees with the maximal inducibility:

Theorem 7.1.0.1 (Theorem 3.6.0.1 in Chapter 3)**.** Let $d \geq 2$ be an arbitrary but fixed positive integer. Among d-ary trees, only binary caterpillars have inducibility 1.

It follows immediately from Theorem 7.1.0.1 that

$$\liminf_{\substack{|T| \to \infty \\ T \ d\text{-ary tree}}} \gamma(D, T) = 0$$

for every d, as soon as D is not a binary caterpillar because

$$0 \leq \liminf_{\substack{|T| \to \infty \\ T \ d\text{-ary tree}}} \gamma(D, T) \leq \liminf_{\substack{|T| \to \infty \\ T \ d\text{-ary tree}}} \left(1 - \gamma(F^2_{|D|}, T)\right) = 1 - I_d\left(F^2_{|D|}\right),$$

provided that D is not isomorphic to $F^2_{|D|}$. However, at this point, it is not clear a priori that

$$\liminf_{\substack{|T| \to \infty \\ T \ d\text{-ary tree}}} \gamma(F^2_k, T) > 0$$

for every k—one will have to put more effort in finding out what the minimum asymptotic density of binary caterpillars might be for every d. Thus, the problem we address in this chapter can be formulated as follows:

Problem: *Given a binary caterpillar F_k^2, is it true that every d-ary tree T with sufficiently large number of leaves always contains a positive density of F_k^2? If so, what is the asymptotic minimum number of copies of F_k^2 in a d-ary tree with large enough number of leaves?*

Let us mention that binary caterpillars have been proved to be extremal among binary trees with respect to some other graph parameters. For instance, binary caterpillars have been shown in [51] to have the maximum Wiener index (sum of distances between all unordered pairs of vertices) among all binary trees with a prescribed number of leaves. In [52], Székely and Wang proved that binary caterpillars minimise the number of subtrees among all binary trees with a given number of leaves.

In the following, we shall prove that the minimum asymptotic density of an arbitrary binary caterpillar is strictly positive for every k. In fact, we shall even be able to derive the precise value of this limiting quantity for every k. Clearly,

$$\liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(F_k^2, T) = 1$$

for $k \leq 2$. Let us now proceed to find its value as a function of d and k for $k \geq 3$.

Our main result reads as follows:

Theorem 7.1.0.2. Let $d, k \geq 2$ be arbitrary but fixed positive integers. Then the following double identity

$$\liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(F_k^2, T) = \liminf_{\substack{|T|\to\infty \\ T\ \text{strictly}\ d\text{-ary tree}}} \gamma(F_k^2, T) = \frac{k!}{2}(d-1)^{k-1}\prod_{j=1}^{k-1}(d^j - 1)^{-1}$$

holds. Furthermore, we have

$$\min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big) \leq \liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma(F_k^2, T)$$

for every k and $n \geq k$.

Formally, the first equality in Theorem 7.1.0.2 tells us that the minimum asymptotic density of any binary caterpillar can also be computed by restricting the set of d-ary trees over which the minimum is taken to strictly d-ary trees only. This situation, in a certain sense, parallels the opposite problem concerning the maximum asymptotic density $I_d(D)$ of a d-ary tree

D, where the authors of paper [34] (see also Chapter 3) could prove that $I_d(D)$ satisfies the equivalent identity

$$I_d(D) = \lim_{n \to \infty} \max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T)$$

for every d-ary tree D. So, it may also be immediately clear that the identity

$$\liminf_{\substack{|T| \to \infty \\ T \ d\text{-ary tree}}} \gamma(F_k^2, T) = \liminf_{\substack{|T| \to \infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma(F_k^2, T)$$

holds for every k. Indeed, it is shown in Theorem 3.3.0.3 (Chapter 3) that for every d-ary tree T with sufficiently large number of leaves, there exists a strictly d-ary tree T^* such that $|T^*| \geq |T|$ and the asymptotic formula

$$\gamma\big(D, T\big) = \gamma\big(D, T^*\big) + \mathcal{O}\big(|T|^{-1}\big) .$$

holds for every d-ary tree D, where the $\mathcal{O}$-constant depends on d only (and nothing else!).

Remark 7.1.0.3. The special cases $d = 2$ and $d = 3$ of Theorem 7.1.0.2 correspond to

$$\liminf_{\substack{|T| \to \infty \\ T \text{ binary tree}}} \gamma\big(F_3^2, T\big) = 1 \quad \text{and} \quad \liminf_{\substack{|T| \to \infty \\ T \text{ ternary tree}}} \gamma\big(F_3^2, T\big) = \frac{3}{4},$$

respectively. In particular, it displays the following equivalence in ternary trees:

$$\liminf_{\substack{|T| \to \infty \\ T \text{ ternary tree}}} \gamma\big(F_3^2, T\big) = 1 - I_3(C_3)$$

as the star C_3 (consisting of a root and three leaves attached to it) and the binary caterpillar F_3^2 are the only 3-leaf d-ary trees for every $d > 2$. Moreover, this confirms that the inducibility of C_3 in ternary trees is $1/4$— see Theorem 3.2.0.1 in Chapter 3.

The next corollary will follow from the proof of Theorem 7.1.0.2:

Corollary 7.1.0.4. Let $d, k \geq 2$ be arbitrary but fixed positive integers. Then the minimum number of copies of the binary caterpillar F_k^2 in an arbitrary n-leaf d-ary tree T is asymptotically

$$\frac{n^k}{2}(d-1)^{k-1}\prod_{j=1}^{k-1}(d^j-1)^{-1} + \mathcal{O}(n^{k-1})$$

as $n \to \infty$.

We should mention that the fact that the binary caterpillar has positive minimum asymptotic density was actually the key result in [1] for the application to the tanglegram crossing problem (see the proof of Lemma 11 in [1]). In fact, using the minimum asymptotic density of binary caterpillars in binary trees, Czabarka, Székely and Wagner [1] proved that the expected value of the number of crossings in a random tanglegram of size n is at least $n^2\big(\frac{2-o(1)}{441}\big)$ with probability at least $1 - n^{-1/2}$ as n tends to infinity.

7.2 Proof of the main theorem and its corollary

This section carries a proof of Theorem 7.1.0.2 as well as a proof of Corollary 7.1.0.4. But before we get to the proofs of these results, we need to go through some preparation.

It is proved in Section 5.3 (Chapter 5) that for a d-ary tree T with branches $T_1, T_2, \ldots, T_d$ (some branches are allowed to be empty), the following recursion

$$c(D,T) = \sum_{i=1}^{d} c(D,T_i) + \sum_{\{i_1,i_2,\ldots,i_r\}\subseteq\{1,2,\ldots,d\}} \ \sum_{\pi\in M(D)} \prod_{j=1}^{r} c\big(D_{\pi(j)}, T_{i_j}\big)\,, \quad (7.2.1)$$

is valid for every d-ary tree T.

Recall that the complete d-ary tree of height h is denoted by CD_h^d.

The following formula can be found explicitly in the proof of Theorem 3.2.0.1 in Chapter 3: for every fixed positive integer $d \geq 2$, we have

$$c(CD_1^r, CD_h^d) = \frac{\binom{d}{r}}{d^r - d}\big(d^{r\cdot h} - d^h\big) \quad (7.2.2)$$

for every $r > 1$ and all $h \geq 1$ (in Chapter 3, the tree CD_1^r is called the r-leaf star).

We also recall that a d-ary caterpillar is a strictly d-ary tree with the property that every internal vertex has $d-1$ adjacent vertices that are leaves, except for the lowest which has d adjacent vertices that are leaves (note that the non-leaf vertices must lie on a single path). We denote the d-ary caterpillar with k leaves by F_k^d.

In the following theorem, we derive an exact formula for the number of copies of the r-ary caterpillar with k leaves in a complete d-ary tree of arbitrary height:

Theorem 7.2.0.1. Let an arbitrary positive integer $d \geq 2$ be fixed. For every $r \in \{2, 3, \ldots, d\}$, the number of copies of F_k^r in CD_h^d is

$$c\big(F_k^r, CD_h^d\big) = \binom{d}{r}^{\frac{k-1}{r-1}} \cdot \Big(\frac{r}{d}\Big)^{\frac{k-r}{r-1}} d^{h-1} \prod_{i=1}^{\frac{k-1}{r-1}} \Big(\frac{d^{h\cdot(r-1)} - d^{(i-1)\cdot(r-1)}}{d^{i\cdot(r-1)} - 1}\Big)$$

for every $k > 1$ and all $h \geq 1$. In particular, we have

$$\lim_{h\to\infty} \gamma(F_k^r, CD_h^d) = \frac{k!}{d}\binom{d}{r}^{\frac{k-1}{r-1}} \Big(\frac{r}{d}\Big)^{\frac{k-r}{r-1}} \prod_{j=1}^{\frac{k-1}{r-1}} \big(d^{(r-1)j} - 1\big)^{-1}.$$

Proof. For $k = r$, the formula of the theorem reads as

$$c\big(F_k^r, CD_h^d\big) = \binom{d}{r} \cdot d^{h-1}\Big(\frac{d^{h\cdot(r-1)} - 1}{d^{r-1} - 1}\Big) = \binom{d}{r}\frac{d^{h\cdot r} - d^h}{d^r - d},$$

and this agrees with equation (7.2.2). For $h = 1$, the formula of the theorem reads as

$$c\big(F_k^r, CD_h^d\big) = \binom{d}{r}^{\frac{k-1}{r-1}} \cdot \Big(\frac{r}{d}\Big)^{\frac{k-r}{r-1}} \prod_{i=1}^{\frac{k-1}{r-1}} \Big(\frac{d^{r-1} - d^{(i-1)\cdot(r-1)}}{d^{i\cdot(r-1)} - 1}\Big),$$

which is equal to 0 as soon as $k > r$, and $\binom{d}{r}$ when $k = r$. So this is again true because for $k > r$, it is clear that there cannot be any copies of F_k^r in CD_1^d. Assume $k > r$ and $h > 1$. It is easy to see that the specialisation $T = CD_h^d$ and $D = F_k^r$ in equation (7.2.1) yields the following recurrence relation:

$$c\big(F_k^r, CD_h^d\big) = d \cdot c\big(F_k^r, CD_{h-1}^d\big) + r \cdot \binom{d}{r} \cdot d^{(h-1)\cdot(r-1)} \cdot c\big(F_{k-r+1}^r, CD_{h-1}^d\big)$$

as all the d branches of CD_h^d are isomorphic to CD_{h-1}^d.

Making use of this recursion, we then continue the proof of the theorem by induction on h. Applying the induction hypothesis, we obtain

$$c(F_k^r, CD_h^d) = \binom{d}{r}^{\frac{k-1}{r-1}} \cdot \left(\frac{r}{d}\right)^{\frac{k-r}{r-1}} \cdot d^{h-1} \left(\prod_{i=1}^{\frac{k-1}{r-1}} \left(\frac{d^{(h-1)\cdot(r-1)} - d^{(i-1)\cdot(r-1)}}{d^{i\cdot(r-1)} - 1} \right) \right.$$
$$\left. + d^{(h-1)\cdot(r-1)} \prod_{i=1}^{\frac{k-r}{r-1}} \left(\frac{d^{(h-1)\cdot(r-1)} - d^{(i-1)\cdot(r-1)}}{d^{i\cdot(r-1)} - 1} \right) \right).$$

We further manipulate this equation and we get

$$\begin{aligned} c(F_k^r, CD_h^d) &= \binom{d}{r}^{\frac{k-1}{r-1}} \cdot \left(\frac{r}{d}\right)^{\frac{k-r}{r-1}} \cdot d^{h-1} \left(d^{1-k} \cdot \prod_{i=1}^{\frac{k-1}{r-1}} \left(\frac{d^{h\cdot(r-1)} - d^{i\cdot(r-1)}}{d^{i\cdot(r-1)} - 1} \right) \right. \\ &\qquad \left. + d^{r-k} \cdot d^{(h-1)\cdot(r-1)} \prod_{i=1}^{\frac{k-r}{r-1}} \left(\frac{d^{h\cdot(r-1)} - d^{i\cdot(r-1)}}{d^{i\cdot(r-1)} - 1} \right) \right) \\ &= \binom{d}{r}^{\frac{k-1}{r-1}} \cdot \left(\frac{r}{d}\right)^{\frac{k-r}{r-1}} \cdot d^{h-1} \\ &\quad \cdot \left(\frac{d^{h\cdot(r-1)} - d^{k-1}}{d^{k-1} \cdot (d^{h\cdot(r-1)} - 1)} + \frac{d^{(h-1)\cdot(r-1)}(d^{k-1} - 1)}{d^{k-r} \cdot (d^{h\cdot(r-1)} - 1)} \right) \\ &\quad \cdot \prod_{i=1}^{\frac{k-1}{r-1}} \left(\frac{d^{h\cdot(r-1)} - d^{(i-1)\cdot(r-1)}}{d^{i\cdot(r-1)} - 1} \right), \end{aligned}$$

completing the induction step and thus the proof of the first part of the theorem. For the assertion on the limit, we note that

$$\begin{aligned} c(F_k^r, CD_h^d) &= \binom{d}{r}^{\frac{k-1}{r-1}} \cdot \left(\frac{r}{d}\right)^{\frac{k-r}{r-1}} \cdot d^{h-1} \left(d^{h\cdot(k-1)} + \mathcal{O}\left(d^{h\cdot(k-r)}\right) \right) \\ &\quad \cdot \prod_{i=1}^{\frac{k-1}{r-1}} \left(d^{i\cdot(r-1)} - 1\right)^{-1} \\ &= \binom{d}{r}^{\frac{k-1}{r-1}} \cdot \left(\frac{r}{d}\right)^{\frac{k-r}{r-1}} \cdot d^{-1} \cdot d^{k\cdot h} \prod_{i=1}^{\frac{k-1}{r-1}} \left(d^{i\cdot(r-1)} - 1\right)^{-1} \\ &\quad + \mathcal{O}\left(d^{h\cdot(k-r+1)}\right), \end{aligned}$$

which implies that

$$\lim_{h\to\infty} \gamma(F_k^r, CD_h^d) = \frac{k!}{d}\binom{d}{r}^{\frac{k-1}{r-1}} \cdot \left(\frac{r}{d}\right)^{\frac{k-r}{r-1}} \prod_{j=1}^{\frac{k-1}{r-1}} \left(d^{(r-1)j} - 1\right)^{-1}$$

as desired. □

Our approach to the proof of Theorem 7.1.0.2 consists of the following steps:

- First, we determine the density of F_k^2 in CD_h^d as $h \to \infty$.
- Next, we prove two auxiliary lemmas.
- Employing the lemmas, we determine an explicit lower bound on $c(F_k^2, T)$ valid for all strictly d-ary trees T.
- Finally, we mention that the bound on $\gamma(F_k^2, T)$ is achieved by complete d-ary trees in the limit.

We replace r with 2 in the formula of $\lim_{h\to\infty} \gamma(F_k^r, CD_h^d)$ given in Theorem 7.2.0.1 to obtain:

Corollary 7.2.0.2. For the k-leaf binary caterpillar F_k^2, we have

$$\lim_{h\to\infty} \gamma(F_k^2, CD_h^d) = \frac{k!}{2}(d-1)^{k-1} \prod_{j=1}^{k-1} (d^j - 1)^{-1}$$

for every $d \geq 2$ and $k \geq 2$.

As a second step, we need two lemmas. Given positive integers $d \geq 2$ and $k \geq 3$, set

$$V_{d,k} = \Big\{ (i_1, i_2, \ldots, i_d) : i_1, i_2, \ldots, i_d \text{ nonnegative integers,} \\ i_1 + i_2 + \cdots + i_d = k, \text{ and none of them is } k \Big\}.$$

Lemma 7.2.0.3. For every given positive integer $k \geq 3$, we have

$$\sup_{\substack{0<x_1,x_2,\ldots,x_d<1 \\ x_1+x_2+\cdots+x_d=1}} \frac{\sum_{1\leq i<j\leq d} \left(x_i \cdot x_j^{-1+k} + x_j \cdot x_i^{-1+k}\right)}{1 - \sum_{i=1}^{d} x_i^k} = \frac{1}{k}$$

for every positive integer $d \geq 2$.

Proof. Fix $d \geq 2$ and $k \geq 3$. Let $V_{d,k}^*$ denote the maximal subset of $V_{d,k}$ that contains no permutation of $\{1, k-1, \underbrace{0, 0, \ldots, 0}_{(d-2)\ 0's}\}$. Then we have the decomposition

$$1 - \sum_{i=1}^{d} x_i^k = \sum_{(i_1,i_2,\ldots,i_d)\in V_{d,k}^*} \binom{k}{i_1, i_2, \ldots, i_d} \prod_{j=1}^{d} x_j^{i_j}$$
$$+ k \cdot \sum_{1\leq i<j\leq d} \left(x_i \cdot x_j^{k-1} + x_i^{k-1} \cdot x_j\right)$$

by the Multinomial Theorem. From that, we immediately deduce the inequality

$$\frac{1}{1 - \sum_{i=1}^{d} x_i^k} \sum_{1\leq i<j\leq d} \left(x_i \cdot x_j^{k-1} + x_i^{k-1} \cdot x_j\right) \leq k^{-1}.$$

This shows that the function

$$F_{d,k}(x_1, x_2, \ldots, x_d) = \frac{\sum_{1\leq i<j\leq d} \left(x_i \cdot x_j^{-1+k} + x_j \cdot x_i^{-1+k}\right)}{1 - \sum_{i=1}^{d} x_i^k}$$

is bounded from above, and so its supremum on the domain defined by $\sum_{i=1}^{d} x_i = 1$ and the inequalities $0 < x_1, x_2, \ldots, x_d < 1$ exists and is finite:

$$\sup_{\substack{0<x_1,x_2,\ldots,x_d<1\\ x_1+x_2+\cdots+x_d=1}} F_{d,k}(x_1, x_2, \ldots, x_d) \leq k^{-1}.$$

On the other hand, we note that

$$F_{d,k}(\underbrace{0, 0, \ldots, 0}_{(d-2)\ 0's}, \epsilon, 1-\epsilon) = \frac{\epsilon \cdot (1-\epsilon)^{k-1} + \epsilon^{k-1} \cdot (1-\epsilon)}{1 - \epsilon^k - (1-\epsilon)^k}$$
$$= \frac{\epsilon \cdot (1-\epsilon)^{k-1} + \epsilon^{k-1} \cdot (1-\epsilon)}{\epsilon\left(\sum_{i=0}^{k-1} (1-\epsilon)^i\right) - \epsilon^k}$$

for every $\epsilon > 0$. It follows that

$$\lim_{\epsilon\to 0} F_{d,k}(\underbrace{0, 0, \ldots, 0}_{(d-2)\ 0's}, \epsilon, 1-\epsilon) = \lim_{\epsilon\to 0} \frac{(1-\epsilon)^{k-1} + \epsilon^{k-2} \cdot (1-\epsilon)}{-\epsilon^{k-1} + \sum_{i=0}^{k-1} (1-\epsilon)^i} = \frac{1}{k},$$

as soon as $k \geq 3$. Hence, we obtain

$$\sup_{\substack{0<x_1,x_2,\ldots,x_d<1\\ x_1+x_2+\cdots+x_d=1}} F_{d,k}(x_1, x_2, \ldots, x_d) = \frac{1}{k},$$

which is the desired result. □

The following lemma gives us the minimum of the function whose supremum is computed in Lemma 7.2.0.3.

Lemma 7.2.0.4. For any positive integers $d \geq 2$ and $k \geq 3$, the function

$$F_{d,k}(x_1, x_2, \ldots, x_d) = \frac{\sum_{1 \leq i < j \leq d} \left(x_i \cdot x_j^{-1+k} + x_j \cdot x_i^{-1+k}\right)}{1 - \sum_{i=1}^{d} x_i^k}$$

subjected to the constraint $\sum_{i=1}^{d} x_i = 1$, on the domain given by the inequalities $0 < x_1, x_2, \ldots, x_d < 1$ has its minimum at $x_1 = x_2 = \cdots = x_d = d^{-1}$, i.e.,

$$F_{d,k}(x_1, x_2, \ldots, x_d) \geq F_{d,k}(\underbrace{d^{-1}, d^{-1}, \ldots, d^{-1}}_{d \text{ terms}}) = \frac{d-1}{d^{k-1}-1}$$

for all $0 < x_1, x_2, \ldots, x_d < 1$ such that $\sum_{i=1}^{d} x_i = 1$.

Proof. Fix $d \geq 2$ and $k \geq 3$. Let S_d be the set of all permutations of the indices $1, 2, \ldots, d$. Since $F_{d,k}(x_1, x_2, \ldots, x_d) > 0$ by definition, the Multinomial Theorem gives

$$\frac{1}{F_{d,k}(x_1, x_2, \ldots, x_d)} = \frac{\sum_{(i_1, i_2, \ldots, i_d) \in V_{d,k}} \binom{k}{i_1, i_2, \ldots, i_d} \prod_{j=1}^{d} x_j^{i_j}}{\sum_{1 \leq i < j \leq d} \left(x_i \cdot x_j^{-1+k} + x_j \cdot x_i^{-1+k}\right)},$$

where $V_{d,k}$ is the set

$$\Big\{(i_1, i_2, \ldots, i_d) : i_1, i_2, \ldots, i_d \text{ nonnegative integers,}$$
$$i_1 + i_2 + \cdots + i_d = k, \text{ and none of them is } k\Big\}.$$

Note that for every $(i_1, i_2, \ldots, i_d) \in V_{d,k}$ such that $i_1 \geq i_2 \geq \cdots \geq i_d$, the vector $(k-1, 1, \underbrace{0, 0, \ldots, 0}_{(d-2)\ 0's})$ majorises $(i_1, i_2, \ldots, i_d)$. Thus, we obtain

$$(d-2)! \cdot \sum_{1 \leq i < j \leq d} \left(x_i \cdot x_j^{-1+k} + x_j \cdot x_i^{-1+k}\right) \geq \sum_{\pi \in S_d} \prod_{j=1}^{d} x_{\pi(j)}^{i_j}$$

by Muirhead's inequality (see the discussion in Section 2.3—Chapter 2). On the other hand, we also have

$$\frac{d!}{F_{d,k}(x_1, x_2, \ldots, x_d)} = \frac{\sum_{(i_1, i_2, \ldots, i_d) \in V_{d,k}} \binom{k}{i_1, i_2, \ldots, i_d} \sum_{\pi \in S_d} \prod_{j=1}^{d} x_{\pi(j)}^{i_j}}{\sum_{1 \leq i < j \leq d} \left(x_i \cdot x_j^{-1+k} + x_j \cdot x_i^{-1+k}\right)}.$$

Therefore, it follows that

$$\frac{d!}{F_{d,k}(x_1,x_2,\ldots,x_d)} \leq (d-2)!\cdot \sum_{(i_1,i_2,\ldots,i_d)\in V_{d,k}} \binom{k}{i_1,i_2,\ldots,i_d} = (d-2)!\cdot(d^k-d)\,,$$

and hence, we establish that

$$F_{d,k}(x_1,x_2,\ldots,x_d) \geq \frac{d-1}{d^{k-1}-1} = F_{d,k}(\underbrace{d^{-1},d^{-1},\ldots,d^{-1}}_{d\text{ terms}})\,.$$

Moreover,

$$F_{d,k}(\underbrace{d^{-1},d^{-1},\ldots,d^{-1}}_{d\text{ terms}}) = \frac{d-1}{d^{k-1}-1}\,.$$

This completes the proof. □

We can now give a proof of Theorem 7.1.0.2.

Proof of Theorem 7.1.0.2. Fix $d \geq 2$. First of all, we want to prove that

$$\liminf_{\substack{|T|\to\infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma(F_k^2, T) = \frac{k!}{2}(d-1)^{k-1}\prod_{j=1}^{k-1}(d^j-1)^{-1} \tag{7.2.3}$$

for every $k \geq 2$. Our approach is an adaptation of [1, Proof of Theorem 7]. Setting

$$b_k = \frac{1}{2}(d-1)^{k-1}\prod_{j=1}^{k-1}(d^j-1)^{-1}\,,$$

we show that for every positive integer $k \geq 2$, the inequality

$$c(F_k^2, T) \geq b_k \cdot n^k - \frac{1}{(k-1)!}\cdot n^{k-1}$$

is satisfied for every strictly d-ary tree T with n leaves.

The case $k = 2$ is essentially obvious as $c(F_2^2, T) = \binom{|T|}{2}$ and $b_2 = 1/2$ by definition. The proof of the general case goes by induction on n. Since $d \geq 2$, it is easy to see that

$$\frac{d^k-1}{d-1} = d^{k-1} + d^{k-2} + \cdots + d + 1 \geq k$$

for every $k \geq 2$. Thus, we have $b_k \leq 1/(2 \cdot (k-1)!) \leq 1/(k-1)!$ meaning that the base case $n = 1$ is true. We can then assume that $k \geq 3$ and $n > 1$. For the induction step, consider the d branches $T_1, T_2, \ldots, T_d$ of an arbitrary strictly d-ary tree T and suppose that they have $\alpha_1 \cdot n$, $\alpha_2 \cdot n, \ldots,$ $\alpha_d \cdot n$ leaves, respectively. In this setting, by replacing D with F_k^2 in equation (7.2.1), we obtain the following formula:

$$c\big(F_k^2, T\big) = \sum_{i=1}^{d} c\big(F_k^2, T_i\big) + \sum_{\substack{1 \leq i,j \leq d \\ i \neq j}} \alpha_i \cdot n \cdot c\big(F_{k-1}^2, T_j\big)\,, \tag{7.2.4}$$

which is valid for every $k \geq 3$. Next, we apply the induction hypothesis: this gives

$$\begin{aligned} c\big(F_k^2, T\big) &\geq \sum_{i=1}^{d} \left(b_k(\alpha_i \cdot n)^k - \frac{1}{(k-1)!}(\alpha_i \cdot n)^{k-1}\right) \\ &\quad + \sum_{\substack{1 \leq i,j \leq d \\ i \neq j}} \alpha_i \cdot n \left(b_{k-1}(\alpha_j \cdot n)^{k-1} - \frac{1}{(k-2)!}(\alpha_j \cdot n)^{k-2}\right) \\ &= \left(b_k \cdot \sum_{i=1}^{d} \alpha_i^k \; + \; b_{k-1} \cdot \sum_{\substack{1 \leq i,j \leq d \\ i \neq j}} \alpha_i \cdot \alpha_j^{k-1}\right) n^k \\ &\quad - \frac{1}{(k-2)!}\left(\frac{1}{k-1} \cdot \sum_{i=1}^{d} \alpha_i^{k-1} + \sum_{\substack{1 \leq i,j \leq d \\ i \neq j}} \alpha_i \cdot \alpha_j^{k-2}\right) n^{k-1}\,. \end{aligned}$$

Using the identity

$$b_k = \frac{d-1}{d^{k-1}-1} \cdot b_{k-1}$$

along with Lemma 7.2.0.4, we get

$$b_k \leq b_{k-1} \cdot \frac{\sum_{\substack{1 \leq i,j \leq d \\ i \neq j}} \alpha_i \cdot \alpha_j^{k-1}}{1 - \sum_{i=1}^{d} \alpha_i^k}\,, \tag{7.2.5}$$

and this takes us to the inequality

$$c\big(F_k^2, T\big) \geq b_k \cdot n^k - \frac{1}{(k-2)!}\left(\frac{1}{k-1} \cdot \sum_{i=1}^{d} \alpha_i^{k-1} + \sum_{\substack{1 \leq i,j \leq d \\ i \neq j}} \alpha_i \cdot \alpha_j^{k-2}\right) n^{k-1}\,.$$

On the other hand, we know from Lemma 7.2.0.3 that

$$\frac{\sum_{\substack{1\le i,j\le d\\ i\neq j}} \alpha_i \cdot \alpha_j^{k-2}}{1-\sum_{i=1}^d \alpha_i^{k-1}} \le \frac{1}{k-1}$$

for all $0 < \alpha_1, \alpha_2, \ldots, \alpha_d < 1$ such that $\sum_{i=1}^d \alpha_i = 1$, provided that $k \ge 4$. In this case, we are done immediately. However, for $k = 3$, equation (7.2.4) becomes

$$\begin{aligned}
c\big(F_3^2, T\big) &= \sum_{i=1}^d c\big(F_3^2, T_i\big) + \sum_{\substack{1\le i,j\le d\\ i\neq j}} \alpha_i \cdot n \cdot \binom{\alpha_j \cdot n}{2} \\
&\ge \sum_{i=1}^d \Big(b_3(\alpha_i \cdot n)^3 - \frac{1}{2}(\alpha_i \cdot n)^2\Big) \\
&\qquad + \frac{1}{2} \cdot \sum_{\substack{1\le i,j\le d\\ i\neq j}} \alpha_i \cdot n \cdot \big((\alpha_j \cdot n)^2 - \alpha_j \cdot n\big) \\
&= \Big(b_3 \cdot \sum_{i=1}^d \alpha_i^3 \; + \; \frac{1}{2} \cdot \sum_{\substack{1\le i,j\le d\\ i\neq j}} \alpha_i \cdot \alpha_j^2\Big) n^3 \\
&\qquad - \frac{1}{2}\Big(\sum_{i=1}^d \alpha_i^2 + \sum_{\substack{1\le i,j\le d\\ i\neq j}} \alpha_i \cdot \alpha_j\Big) n^2 \,,
\end{aligned}$$

where the inequality in the second step follows from the induction hypothesis. Therefore, using the identity

$$1 - \sum_{i=1}^d \alpha_i^2 = \sum_{\substack{1\le i,j\le d\\ i\neq j}} \alpha_i \cdot \alpha_j \,,$$

together with inequality (7.2.5) (as $b_2 = 1/2$), we deduce that

$$c\big(F_3^2, T\big) \ge b_3 \cdot n^3 - \frac{1}{2} \cdot n^2 \,,$$

and this completes the induction proof. Notice that the right side of equation (7.2.3) appears already in Corollary 7.2.0.2. That is, we have

$$\liminf_{\substack{|T|\to\infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma\big(F_k^2, T\big) = \lim_{h\to\infty} \gamma\big(F_k^2, CD_h^d\big) \,,$$

and this finishes the proof of the first assertion of Theorem 7.1.0.2.

Let us now tackle the second part of the theorem. The proof is similar to that of Theorem 3.3.0.1 in Chapter 3.

Claim: The sequence

$$\Big(\min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big)\Big)_{n\geq k}$$

is nondecreasing, that is, we have

$$\min_{\substack{|T|=n-1 \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big) \leq \min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big)$$

for every $n \geq 1 + k$.

For the proof of the claim, let T be a d-ary tree with leaf-set $L(T)$ such that $|T| \geq k$. For $l \in L(T)$, denote by $c_l(F_k^2, T)$ the number of subsets of leaves of T that involve l and induce a copy of F_k^2. Thus, every leaf of T is involved in $k \cdot c(F_k^2, T)/|T|$ copies of F_k^2 on average, and so there exists a leaf l_1 of T for which the inequality

$$c_{l_1}(F_k^2, T) \geq \frac{k \cdot c(F_k^2, T)}{|T|} \tag{7.2.6}$$

holds. The number of copies of F_k^2 in T not involving the leaf l_1 is

$$c(F_k^2, T) - c_{l_1}(F_k^2, T) \leq \Big(1 - \frac{k}{|T|}\Big) c(F_k^2, T)$$

by virtue of relation (7.2.6). Call T^- the d-ary tree that results when the leaf l_1 of T is removed and the unique vertex adjacent to l_1 (if it has outdegree 2 in T) is suppressed. Thus, since T is an arbitrary d-ary tree, we get

$$\min_{\substack{|T'|=n-1 \\ T'\ d\text{-ary tree}}} c(F_k^2, T') \leq c(F_k^2, T^-) \leq \Big(1 - \frac{k}{n}\Big) \min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} c(F_k^2, T)\,,$$

so that dividing both sides of this inequality by $\binom{n-1}{k}$, we obtain

$$\min_{\substack{|T'|=n-1 \\ T'\ d\text{-ary tree}}} \gamma\big(F_k^2, T'\big) \leq \min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big)$$

for every $n \geq 1 + k$, showing that the sequence

$$\left(\min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma(F_k^2, T) \right)_{n \geq k}$$

is indeed nondecreasing. Hence, since this sequence is also bounded from above for every $n \geq k$, one obtains

$$\lim_{n\to\infty} \min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big) = \liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big)$$

and consequently, we get

$$\min_{\substack{|T|=n \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big) \leq \liminf_{\substack{|T|\to\infty \\ T\ d\text{-ary tree}}} \gamma\big(F_k^2, T\big)$$

for every $n \geq k$. This completes the entire proof of the theorem. □

The proof of Corollary 7.1.0.4 is now immediate as

$$c\big(F_k^2, CD_h^d\big) = \frac{(d-1)^{k-1}}{2} \cdot d^h \prod_{i=1}^{k-1} \left(\frac{d^h - d^{i-1}}{d^i - 1} \right)$$

for all $h \geq 1$ (see Theorem 7.2.0.1).

7.3 Conclusion

We conclude this short chapter with an open question. To formalise the question, we need to define a new class of binary trees. These trees are already considered in Chapter 5 and [1]. A binary tree T is called even if for every internal vertex v of T, the number of leaves in the two branches of the subtree of T rooted at v differ at most by one. It is easy to see that there is only one such a binary tree for every given number of leaves. We denote the n-leaf even binary tree by E_n^2.

QUESTION 7.3.0.1. Is it true that for $n \geq k$, the even binary tree E_n^2 has the smallest number of copies of the binary caterpillar F_k^2 among all binary trees with n leaves?

We mention that the case $k \leq 3$ is trivial, while calculations show that the case $k \in \{4, 5\}$ is also true for values of n up to 100.

Bibliography

[1] Czabarka, É., Székely, L.A. and Wagner, S.: Inducibility in binary trees and crossings in random tanglegrams. *SIAM Journal on Discrete Mathematics*, vol. 31, no. 3, pp. 1732–1750, 2017.

[2] Pippenger, N. and Golumbic, M.C.: The inducibility of graphs. *Journal of Combinatorial Theory, Series B*, vol. 19, no. 3, pp. 189–203, 1975.

[3] Bollobás, B., Nara, C. and Tachibana, S.-i.: The maximal number of induced complete bipartite graphs. *Discrete mathematics*, vol. 62, no. 3, pp. 271–275, 1986.

[4] Exoo, G.: Dense packings of induced subgraphs. *Ars Combinatoria*, vol. 22, pp. 5–10, 1986.

[5] Brown, J.I. and Sidorenko, A.: The inducibility of complete bipartite graphs. *Journal of Graph Theory*, vol. 18, no. 6, pp. 629–645, 1994.

[6] Bollobás, B., Egawa, Y., Harris, A. and Jin, G.: The maximal number of induced r-partite subgraphs. *Graphs and Combinatorics*, vol. 11, no. 1, pp. 1–19, 1995.

[7] Hirst, J.: The inducibility of graphs on four vertices. *Journal of Graph Theory*, vol. 75, no. 3, pp. 231–243, 2014.

[8] Razborov, A.A.: Flag algebras. *The Journal of Symbolic Logic*, vol. 72, no. 04, pp. 1239–1282, 2007.

[9] Razborov, A.A.: Flag algebras: an interim report. In: *The Mathematics of Paul Erdős II*, pp. 207–232. Springer, 2013.

[10] Razborov, A.A.: What is a flag algebra. *Notices of the AMS*, vol. 60, no. 10, pp. 1324–1327, 2013.

[11] Grzesik, A.: *Flag algebras in extremal graph theory*. Ph.D. thesis, PhD thesis, Jagiellonian University, Krakow, 2014.

[12] Razborov, A.A.: On 3-hypergraphs with forbidden 4-vertex configurations. *SIAM Journal on Discrete Mathematics*, vol. 24, no. 3, pp. 946–963, 2010.

[13] Hladkỳ, J., Norin, S. *et al.*: Counting flags in triangle-free digraphs. *Combinatorica*, vol. 37, no. 1, pp. 49–76, 2017.

[14] Lovász, L.: Semidefinite programs and combinatorial optimization. In: *Recent advances in algorithms and combinatorics*, pp. 137–194. Springer, 2003.

[15] Hatami, H., Hirst, J. and Norine, S.: The inducibility of blow-up graphs. *Journal of Combinatorial Theory, Series B*, vol. 109, pp. 196–212, 2014.

[16] Even-Zohar, C. and Linial, N.: A note on the inducibility of 4-vertex graphs. *Graphs and Combinatorics*, vol. 31, no. 5, pp. 1367–1380, 2015.

[17] Balogh, J., Hu, P., Lidickỳ, B. and Pfender, F.: Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle. *European Journal of Combinatorics*, vol. 52, pp. 47–58, 2016.

[18] Sperfeld, K.: The inducibility of small oriented graphs. *arXiv preprint arXiv:1111.4813*, 2011.

[19] Huang, H.: On the maximum induced density of directed stars and related problems. *SIAM Journal on Discrete Mathematics*, vol. 28, no. 1, pp. 92–98, 2014.

[20] Chor, B. and Steel, M.: Do tree split probabilities determine the branch lengths? *Journal of theoretical biology*, vol. 374, pp. 54–59, 2015.

[21] Sedgewick, R. and Flajolet, P.: *An introduction to the analysis of algorithms*. Addison-Wesley, 2013.

[22] Aho, A.V. and Hopcroft, J.E.: *The design and analysis of computer algorithms*. Pearson Education India, 1974.

[23] Butcher, J.C.: *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*. Wiley-Interscience, 1987.

[24] Butcher, J.C.: *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.

[25] Wiener, H.: Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, vol. 69, no. 1, pp. 17–20, 1947.

[26] Cyvin, S.J. and Gutman, I.: *Kekulé structures in benzenoid hydrocarbons*, vol. 46. Springer Science & Business Media, 2013.

[27] Rosen, K.H.: Discrete mathematics and its applications. *AMC*, vol. 10, p. 12, 2007.

[28] Song, Y.S.: On the combinatorics of rooted binary phylogenetic trees. *Annals of Combinatorics*, vol. 7, no. 3, pp. 365–379, 2003.

[29] Bubeck, S. and Linial, N.: On the local profiles of trees. *Journal of Graph Theory*, vol. 81, no. 2, pp. 109–119, 2016.

[30] Trivers, R. and Burt, A.: Genes in conflict: The biology of selfish genetic elements. 2006.

[31] Hafner, M.S. and Nadler, S.A.: Phylogenetic trees support the coevolution of parasites and their hosts. *Nature*, vol. 332, no. 6161, p. 258, 1988.

[32] Venkatachalam, B., Apple, J., John, K.S. and Gusfield, D.: Untangling tanglegrams: Comparing trees by their drawings. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 7, no. 4, pp. 588–597, 2010.

[33] Semple, C. and Steel, M.A.: *Phylogenetics*, vol. 24. Oxford University Press on Demand, 2003.

[34] Czabarka, E., Dossou-Olory, A.A.V., Székely, L.A. and Wagner, S.: Inducibility of d-ary trees. *arXiv preprint arXiv:1802.03817*, 2018.

[35] Dossou-Olory, A.A.V. and Wagner, S.: Inducibility of topological trees. *Accepted for publication in Quaestiones Mathematicae*, 2018.

[36] Dossou-Olory, A.A.V.: The minimum asymptotic density of binary caterpillars. *arXiv preprint arXiv:1804.05731*, 2018.

[37] Dossou-Olory, A.A.V. and Wagner, S.: Further results on the inducibility of d-ary trees. *To be submitted*, 2018.

[38] Dossou-Olory, A.A.V. and Wagner, S.: On the inducibility of small trees. *To be submitted*, 2018.

[39] West, D.B. *et al.*: *Introduction to Graph Theory*, vol. 2. Prentice hall Upper Saddle River, 2001.

[40] Harary, F. and Palmer, E.M.: *Graphical enumeration*. Elsevier, 2014.

[41] Bergeron, F., Labelle, G. and Leroux, P.: *Combinatorial species and tree-like structures*, vol. 67. Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1998.

[42] Allman, E.S. and Rhodes, J.A.: *Mathematical models in biology: an introduction*. Cambridge University Press, 2004.

[43] Genitrini, A.: Full asymptotic expansion for polya structures. *arXiv preprint arXiv:1605.00837*, 2016.

[44] The on-line encyclopedia of integer sequences, published electronically at `https://oeis.org`. 2018.

[45] Carothers, N.L.: *Real analysis*. Cambridge University Press, 2000.

[46] Hata, M.: *Problems and solutions in real analysis*, vol. 4. World Scientific Publishing Co Inc, 2007.

[47] Hardy, G.H., Littlewood, J.E. and Pólya, G.: *Inequalities*. Cambridge university press, 1952.

[48] Prodinger, H. and Tichy, R.: Fibonacci numbers of graphs. *The Fibonacci Quarterly*, vol. 20, no. 1, pp. 16–21, 1982.

[49] Gutman, I.: Acyclic systems with extremal hückel π-electron energy. *Theoretica chimica acta*, vol. 45, no. 2, pp. 79–87, 1977.

[50] Gutman, I. and Polansky, O.E.: Mathematical concepts in organic chemistry. 1986.

[51] Fischermann, M., Hoffmann, A., Rautenbach, D., Székely, L. and Volkmann, L.: Wiener index versus maximum degree in trees. *Discrete Applied Mathematics*, vol. 122, no. 1-3, pp. 127–137, 2002.

[52] Székely, L.A. and Wang, H.: On subtrees of trees. *Advances in Applied Mathematics*, vol. 34, no. 1, pp. 138–155, 2005.